

DIBELS Next®: Findings from the Benchmark Goals Study

Kelly A. Powell-Smith

Roland H. Good

Rachael J. Latimer

Elizabeth N. Dewey

Joshua Wallin

Ruth A. Kaminski

Dynamic Measurement Group

Technical Report # 11

Suggested Citation

Powell-Smith, K.A., Good, R.H., Latimer, R. J., Dewey, E. N., Wallin, J., & Kaminski, R.A.(2012). *DIBELS Next®: Findings from the Benchmark Goals Study* (Technical Report No. #11). Eugene, OR: Dynamic Measurement Group.

Author Note

The authors thank the faculty, students, parents, and school personnel for their effort and support during the course of this study. We also thank Dr. Courtney E. Wheeler for her contributions during the writing and editing of this report. Correspondence regarding this manuscript should be addressed to Kelly A. Powell-Smith, Dynamic Measurement Group, 859 Willamette St., Suite 320, Eugene, OR 97401; Email: kpowellsmith@dibels.org.

DIBELS Next: Findings from the Benchmark Goals Study

The purpose of this technical report is to detail the findings from a study designed to evaluate the measures that comprise DIBELS Next as well as describe the process of determining new benchmark goals and cut points for risk. In this report we provide information about the development of the measures under review, along with descriptive and correlational data. For further information regarding the development of these measures, as well as additional technical reports, please visit: www.dibels.org.

Method

Participants

The participants in this study were students in grades kindergarten through sixth grade in 13 elementary schools from 5 school districts in five states across three regions in the United States. The data were collected during the 2009-2010 school year. The students were selected from general education classrooms who were receiving English language reading instruction, including students with disabilities who are English language learners, provided they had the response capabilities to participate.

To examine the reliability and validity of the DIBELS measures, further testing was performed with subsets of the total sample. For validity testing, one subset of 50 students from each of the five participating sites received the GRADE for grades K through 6. For reliability testing, three of the five sites conducted one of the following: test-retest reliability, alternate-form reliability, or inter-rater reliability testing.

Examiners, teachers, and administrators who participated in administration and scoring of DIBELS Next at the sites were asked to complete a Consumer Satisfaction Questionnaire (BGS Usability Questionnaire) electronically via a web-based tool called Survey Monkey. To preserve

anonymity, limited demographic data were collected on the personnel (e.g., teachers, assessors, and administrators) who completed the BGS Usability Questionnaire. Respondents were asked to indicate their school district and their role in the school district. No further demographic data were collected. Of the 138 personnel who completed the usability questionnaire, 69% (n = 95) identified themselves as teachers. Seventeen percent (n = 24) identified themselves as both teachers and assessors. Ten percent (n = 14) identified themselves as assessors only. Finally, 4% (n = 5) of the sample were administrators.

Demographic information-school level. The schools involved in the Benchmark Goals Study represent primarily the Midwest portion of the United States (see Table 1). Experience using DIBELS within the participating districts ranged from four to ten years.

Table 1
Research Site Information

Research site	Census Region, Division	State	DIBELS® Experience (years)
A	Midwest, East North Central	OH	10
B	Midwest, West North Central	IA	10
C	West, Pacific	CA	4
D	Midwest, West North Central	MN	10
E	West, Pacific	OR	9
Median :			10

Note. The years of DIBELS® experience is at the district level, and may not reflect the number of years of experience of all schools within a district.

Data for this study were captured, primarily, through an extant database (i.e., the University of Oregon Data System, <https://dibels.uoregon.edu/>). As a result, some demographic information was not reported at the student level. Demographic data at the school level were gathered from

the National Center for Education Statistics (NCES) website (<http://nces.ed.gov/>) for the 2008-2009 school year. Tables 2 and 3 summarize these data. Data on the number of ELL students, students with IEP's, and expenditure per student were summarized at the district level and therefore include information on students who were outside the age ranges included in the study (e.g., pre-K, and grades 7 -12). The total number of students in each district, only including participating schools, ranged from 255 students in District 5 to 1682 students in District 2, with a student to teacher ratio ranging from 10.3 in District 3 to 18.0 in District 1. The student populations at the participating schools were primarily Caucasian, with percentages ranging from 85% in District 3 to 98% in District 1. Schools in District 3 reported the highest minority population with 5% of the population identifying as Asian, and 8% of the population identifying as Hispanic.

Table 2

United States and Research Site Demographic Comparisons

Population	Total schools	Total students	Student : Teacher ratio	Expenditure per student
District 1	2	806	18.0	\$9,428
District 2	3	1682	12.9	\$9,272
District 3	1	571	10.3	\$16,182
District 4	5	1278	16.9	\$10,562
District 5	1	255	17.2	\$3,027
U.S. Primary & Secondary schools	132,436	49,298,945	15.8	\$10,041

Population	Total schools	ELL students	Students with IEPs	Free/Reduced lunch eligible
District 1	2	2	135	0
District 2	3	9	310	300
District 3	1	34	51	20
District 4	5	45	265	302
District 5	1	15	82	96

Note. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data (CCD) for the 2008-09 school year. Fiscal data available for the 2007-08 school year. Data is based on actual reported numbers and may not include students who elected to not report these data. District 4 includes data for 2 schools from the PSS Private School Universe Survey for the 2007-08 school year. "N/A" indicates the data are not available or not applicable. English Language Learners (ELLs), students with IEPs (Individualized Education Programs), and expenditure per student information is reported at the district level as it is unavailable at the school level, and therefore may include grades not involved in the study, such as pre-K, and grades 7 through 12. Districts 1, 2, and 4 include grades not involved in the study, such as pre-K, 7, and/or 8. "United States Primary and Secondary" totals represent data from the 2005-06 school year. All schools were Title 1 eligible, with the exception of one school in District 2 and 3, and two schools in District 4.

Table 3

Demographic Information by Site Compared with Total U.S. Population

Population	Race/Ethnicity					Total
	American Indian or Alaska Native	Asian	Black or African American	Hispanic	White	
District 1	0	< 1% (1)	1% (6)	1% (5)	98% (784)	796
District 2	< 1% (2)	1% (11)	1% (15)	2% (32)	96% (1622)	1682
District 3	1% (3)	5% (31)	1% (5)	8% (47)	85% (478)	564
District 4	< 1% (4)	1% (7)	1% (13)	7% (89)	91% (1149)	1262
District 5	< 1% (1)	0	2% (4)	4% (10)	91% (232)	247
Total	< 1% (6)	1% (53)	< 1% (38)	4% (183)	94% (4347)	4627
US population under 18 years	1% (840,000)	3% (2.5 million)	14% (10.9 million)	16% (12.3 million)	65% (49.6 million)	76.14 million

Note. All data are reported from the National Center for Education Statistics (NCES) for the 2008-2009 school year. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Population data are the aggregate of school level information. Districts 1, 2, and 4 include grades not involved in the study, such as pre-K, 7, and/or 8. Data for the total U.S. population under 18 years are from the 2000 Census.

Although not traditionally considered demographic information, the initial skill level of students in the fall is a strong determinant of their gains over the course of the school year (Hedges & Hedberg, 2007; Bloom, Richburg-Hayes, and Black, 2007; and Slavin, 2008) and the best way to at least partially equate students and schools. No other predictors routinely collected, especially demographic information, are as good at explaining both student and school level

differences. This information is reported in Table 16 in the Results section of this technical report.

A subsample of the students was selected to participate in GRADE testing. The caregivers of the students participating in GRADE testing were asked to complete a demographics survey about their student and household. This student- and household-level information provides a better description of the student population that participated in this portion of the study. Approximately 97% of students involved in GRADE testing returned the additional demographic surveys. This information is presented in Tables 4 and 5. This subsample of students reported similar racial and ethnic background percentages as that reported to NCES. Of the returned demographic surveys, the student population reported about 50% female, and a range from 88% in District 3 to 97% in District 1 identifying as Not Hispanic. As reported, the largest percentage overall for the highest level of education of a parent or guardian was a 4-year college education, although there was variability across districts. The greatest percentage for parent or guardian education in District 5 was 40% at the High School level, and the greatest percentage reported in District 3 was 36% at the Master's degree level. Total household income levels also varied from 20% in District 1 reporting a household income level of \$14,570 or less, and 90% of District 3 households reporting an income of over \$100,000. Overall, the largest percentages of household incomes were over \$50,000.

Table 4

Parent-Reported Demographic Information for Students Receiving the GRADE

Student Demographic Category	Population					
	District 1	District 2	District 3	District 4	District 5	Total
Gender						
Female	52% (86)	49% (112)	47% (75)	53% (175)	47% (94)	50% (542)
Male	48% (80)	51% (118)	53% (84)	47% (154)	53% (107)	50% (543)
Ethnic Background						
Hispanic or Latino/a	0% (0)	5% (12)	11% (17)	7% (23)	6% (12)	6% (64)
Not Hispanic or Latino/a	97% (142)	92% (208)	88% (139)	92% (290)	94% (176)	92% (955)
Other	3% (5)	2% (5)	1% (2)	1% (2)	0% (0)	1% (14)
Racial Background						
American Indian or Native Alaskan	1% (2)	1% (2)	0% (0)	1% (2)	4% (7)	1% (13)
Asian	0% (0)	1% (2)	3% (5)	0% (0)	1% (1)	< 1% (8)
Black or African American	1% (1)	< 1% (1)	2% (3)	0% (0)	1% (2)	< 1% (7)
Multiracial	2% (4)	5% (11)	8% (13)	2% (5)	9% (17)	5% (50)
Native Hawaiian or Other Pacific Islander	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
White	96% (159)	92% (212)	85% (131)	93% (302)	84% (168)	91% (972)
Other	0% (0)	1% (2)	2% (3)	4% (12)	3% (5)	2% (22)
Unknown	0% (0)	0% (0)	0% (0)	1% (2)	0% (0)	< 1% (2)

Note. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Percent of students that reported any demographic information is 97% ($n = 1240$).

Table 5

Parent-Reported Demographic Information for Households of Students Receiving the GRADE

Household Demographic Category	Population					
	District 1	District 2	District 3	District 4	District 5	Total
Highest Level of Education						
Grade School	2% (3)	0% (0)	0% (0)	2% (5)	0% (0)	1% (8)
Middle School/Junior High School	2% (4)	0% (0)	3% (5)	3% (8)	0% (0)	2% (17)
High School	35% (57)	10% (22)	1% (2)	11% (33)	40% (75)	18% (189)
2-year college	13% (21)	22% (49)	4% (7)	19% (59)	16% (29)	16% (165)
4-year college	15% (24)	33% (75)	31% (49)	28% (88)	13% (25)	25% (261)
Vocational or Technical Training	18% (29)	9% (21)	3% (5)	18% (55)	17% (31)	13% (141)
Some graduate training	4% (6)	10% (22)	9% (14)	4% (12)	5% (9)	6% (63)
Completed Masters degree	12% (19)	12% (28)	36% (56)	15% (46)	9% (16)	16% (165)
Completed Doctoral degree	1% (1)	4% (8)	12% (19)	3% (8)	1% (1)	4% (37)
Total Household Income						
\$14,570 or less	20% (29)	1% (3)	0% (0)	5% (14)	8% (14)	6% (60)
\$14,571 - \$18,310	7% (11)	0% (0)	1% (1)	1% (3)	2% (3)	2% (18)
\$18,311 - \$22,050	5% (7)	1% (3)	1% (1)	3% (9)	2% (4)	3% (24)
\$22,051 - \$25,790	1% (2)	2% (5)	3% (4)	1% (3)	6% (10)	3% (24)
\$25,791 - \$29,530	3% (4)	< 1% (1)	1% (2)	3% (8)	3% (6)	2% (21)
\$29,531 - \$33, 270	1% (2)	3% (6)	1% (1)	8% (21)	2% (3)	3% (33)
\$33,271 - \$37,010	8% (12)	2% (5)	0% (0)	4% (10)	5% (8)	4% (35)
\$37,011 - \$49,999	9% (13)	8% (18)	0% (0)	10% (27)	17% (29)	9% (87)

\$50,000 - \$74,999	20% (30)	21% (45)	1% (2)	25% (68)	30% (52)	21% (197)
\$75,000 - \$99,999	15% (22)	21% (46)	2% (3)	17% (46)	17% (30)	16% (149)
\$100,000 or more	10% (15)	39% (83)	90% (132)	21% (56)	6% (10)	32% (300)

Note. Data is based on actual reported numbers, indicated in parentheses, and may not include students who elected to not report these data. Percent of students that reported any demographic information is 97% ($n = 1240$).

Procedures

All *DIBELS Next* testing was conducted by onsite coordinators, school teachers, or other school personnel who were already familiar with the DIBELS measures and routinely collect DIBELS data with the 6th Edition DIBELS measures. GRADE testing also was conducted by onsite coordinators, school teachers and other school personnel who participated in training on this assessment. Details of training follow.

DIBELS Next training. Data collectors participated in a one-day training conducted by DMG research scientists or professional development staff. A review of the materials and administration and scoring rules was provided during this time. Examiners were provided with opportunities for practice and to ask questions during the training. During the training, examiners were given sample items from each measure, except Daze, to practice correct scoring procedures. Agreement on scoring and fidelity to standardize administration and scoring procedures was checked.

In addition to this one-day training, all onsite coordinators participated in score calibration activities at the beginning of the study (except for one site where the coordinator was shadow-scored by DMG personnel already deemed to meet reliability criteria for administration & scoring of DIBELS Next). These score calibration activities were designed to ensure that coordinators were reliable and accurate in their scoring prior to them either testing students or training other data collectors. Calibration consisted of examiners listening to and scoring a DIBELS measure while the principal investigator role-played the part of examiner and student. Each role play followed a pre-specified script. This process was completed for each DIBELS measure except DAZE. All calibration activities were conducted individually with each site coordinator via conference call. Results of this calibration testing are reported in Table 6.

Table 6. Percent Scoring Accuracy Results by Measure and Site Coordinator

Measure	1	Site Coordinator Number			4	Average Accuracy
		2a	2b	3		
LNF	100%	100%	100%	97%	100%	99%
FSF	100%	100%	100%	100%	100%	100%
PSF	100%	97%	97%	100%	100%	99%
NWF	100%	100%	100%	100%	99%	100%
ORF	100%	100%	100%	100%	100%	100%
Retell	98%	96%	100%	100%	95%*	98%
WUF-R	96%	84%	100%	96%	96%	94%
Average Accuracy	99%	97%	100%	99%	99%	99%

Note. LNF = Letter Naming Fluency; FSF = First Sound Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; WUF –R = Word Use Fluency – Revised. * on 2nd attempt.

In addition to scoring accuracy, whether site coordinators provided DIBELS reminders and prompts during the testing was observed. It should be noted that the coordinators were at a disadvantage because they were not technically in the administrator role. Therefore, providing reminders to someone serving as both student and examiner presented a somewhat awkward situation. Although these behaviors were tracked and doing so provided the means to discuss the use or reminders during testing, data obtained are likely an underestimate of examiners' procedural accuracy. Any discrepancies in scoring or procedures (e.g., use of prompts and reminders) were discussed prior to concluding the calibration session. Detailed notes on the performance of each site coordinator are described in Figure 1.

Summary Notes of Calibration Observations for Site Coordinators
Site Coordinator #1 provided reminders as indicated on all measures except for FSF. The missed reminder was <i>"Remember to tell me the first sound in the word, not the letter name."</i> It is noteworthy that the response that should have elicited this reminder from the examiner occurred near the end of the one-minute timing. Site Coordinator #2a provided reminders and prompts as indicated on all DIBELS Next measures. On NWF there was a slight delay in providing the three second hesitation prompt and on Retell the "thank you" to be given after the second hesitation for five seconds was slightly premature. Site Coordinator #2b provided the reminders and prompts for LNF and NWF. She did not provide the three second hesitation reminder on Retell or give the <i>"Remember to say <u>all</u> the sounds in the word"</i> reminder for PSF. In addition, she did not provide the reminder for FSF, <i>"Remember to tell me the first sound in the word, not the letter name."</i> It is noteworthy that the response that should have elicited this reminder from the examiner occurred near the end of the one-minute timing. Finally, she did not provide the <i>"Remember to say a sentence with the word"</i> prompt for WUF-R. Site Coordinator #3 provided the correct reminders and prompts for ORF, Retell, and WUF-R. She also provided reminders on LNF & NWF, but using non-standard language. On PSF she did not provide the reminder, <i>"Remember to say <u>all</u> the sounds in the word."</i> She did not provide the reminder for FSF. The missed reminder was <i>"Remember to tell me the first sound in the word, not the letter name."</i> It is noteworthy that the response that should have elicited this reminder from the examiner occurred near the end of the one-minute timing. Site Coordinator #4 provided the correct reminders and prompts for LNF, NWF, and ORF. On Retell, she did not provide the three second hesitation reminder. In addition, she did not provide the <i>"Remember to say a sentence with the word"</i> prompt for WUF-R and did not provide the FSF reminder to <i>"Remember to tell me the first sound in the word, not the letter name."</i> Finally, on PSF she did not provide the reminder, <i>"Remember to say <u>all</u> the sounds in the word."</i>

Figure 1. Notes on Coordinator Administration Calibration

GRADE training. Training in the administration of the GRADE was conducted via webcast. Three webcasts were conducted and each was approximately 2 hours in length. Each webcast had two to three participants. Site Coordinators and other key school personnel participated in these webcasts. Participants were provided with handout versions of PowerPoint slides used during the training, fidelity checklists for each level of the GRADE test, and a table summarizing

information about GRADE subtests, a description of those subtests, the levels at which they are given, and the composite score(s) to which each subtest contributes.

During the training, the importance of the role of data collector was reviewed first. Next, an introduction to the GRADE was provided including a discussion of: (a) who can administer the GRADE, (b) the materials used for test administration, and (c) the various GRADE subtests including the stimulus and response demands for each subtest and the grade levels at which they are given. Then, GRADE administration procedures were reviewed. Emphasis was placed on adherence to standardized procedures. Next, each level of the GRADE used in this study was reviewed with participants. The number of test sessions, the approximate length of each session, and general administration procedures for each subtest were discussed. Every GRADE subtest was practiced by participants in a round-robin manner. Feedback was provided. Excellent adherence to standardized procedures and verbatim directions was noted. Finally, the logistics of GRADE administration were discussed including tips for ensuring the GRADE data collection process runs smoothly and what to do if problems are encountered.

Inter-Rater reliability. Each site was asked to conduct shadow-scoring of *DIBELS Next* testing with 5 students per grade in grades K – 6. They were asked to select students at random from grade-level lists and, if possible, to have every examiner participate as both assessor and shadow-scorer. In addition, sites were asked to make sure student names were not on the shadow scored protocols. Instead, these protocols were labeled with student ID, grade, and district. Any student names were removed from shadow-scoring protocols and returned to DMG. This process was used for all measures except for Daze. For Daze, each site randomly selected 5 students in grades 3 - 6 (grades 3 - 5 for District E) and copied the Daze worksheet for each of those students. These worksheets were mailed back to DMG and then re-scored by a DMG research

assistant. Sites were asked to return Daze worksheets without student names. However, one site sent their Daze protocols with student names. These were destroyed (shredded) and another set of 5 Daze worksheets per grade level were selected, sent to DMG, and re-scored.

Alternate form reliability. Alternate-form reliability testing was conducted for *DIBELS Next* measures two weeks after middle of year DIBELS benchmark assessment. All testing was conducted at a single site, District C. This site was chosen because it reported a greater variability in student reading skill. Students were sampled by principal investigator using the fall benchmark data. Stratified sampling by benchmark status was utilized to obtain a sample comprised of 50% students at benchmark and 50% from combined strategic and intensive instructional recommendation categories. First, each grade level list was sorted by instructional recommendation. Second, 20 students with benchmark instructional recommendations were randomly selected using the random number function in Excel. Third, 13 students were selected randomly from each of the strategic and intensive instructional recommendation levels, again using the random number function in Excel. This procedure was used at each grade level with the exception of first grade and fifth grade. In first grade, all of the students with strategic and intensive instructional recommendations were selected for the consenting process due to low numbers ($n = 14$ strategic, $n = 6$ intensive). In fifth grade, due to a disproportionate number of students with strategic (12 students) and intensive (29 students) instructional recommendations, all 12 strategic students and 14 of the 29 intensive students were selected for the consenting process to bring the proportion of students below benchmark to 50%.

The sampling goal was 30 students per grade (i.e., 15 benchmark and 15 from strategic/intensive combined). To ensure the sampling goal was met, oversampling of some instructional recommendation categories in some grade levels (i.e., grade 5 intensive) was

conducted even though the target numbers to test were the same across the strategic and intensive groups.

The sample size breakdown by grade level and instructional recommendation (based on fall data) for students who had consent forms sent home is reported in Table 7. The data in Table 7 indicates that our sampling goal was met, for the most part. However, in kindergarten, third, fourth and fifth grades there were more than 50% of students performing below benchmark (i.e., at strategic or intensive levels).

Table 7

Alternate Form Sample Stratification

Status by grade	N	Consents	Ideal	Actual	Percentage
Kindergarten					
Benchmark	35	20	15	13	45
Strategic	25	13	8	9	31
Intensive	25	13	8	7	24
First Grade					
Benchmark	44	26	15	15	54
Strategic	14	14	10	10	36
Intensive	6	6	6	3	11
Second Grade					
Benchmark	54	20	15	12	50
Strategic	24	13	8	5	21
Intensive	18	13	8	7	29
Third Grade					
Benchmark	34	20	15	14	44

Strategic	23	13	8	10	30
Intensive	25	13	8	9	27
Fourth Grade					
Benchmark	27	20	15	14	41
Strategic	21	13	8	9	26
Intensive	45	13	8	11	32
Fifth Grade					
Benchmark	47	20	15	12	48
Strategic	12	12	8	7	28
Intensive	29	14	8	6	24
Sixth Grade					
Benchmark	61	20	15	7	50
Strategic	21	13	8	6	43
Intensive	13	13	8	1	7

Note. District A data only. Stratified sampling goal was 50% of the sample from Benchmark status, and 50% of the sample evenly divided between Strategic and Intensive. N = sample from which student ID numbers were drawn to receive consent forms; Consents = number of students that received consent forms; Ideal = the ideal number of students that send consent forms back; Actual = the actual number of students that sent consent forms back; Percentage = proportion of students that returned consent forms from each group.

Test-Retest reliability. Retesting with *DIBELS Next* measures was conducted two weeks after the middle of year DIBELS benchmark assessment. All testing was conducted at a single site, District C; which volunteered to conduct test-retest reliability. Students were sampled by the principal investigator based upon data from the fall benchmark. Stratified sampling by benchmark status was utilized to obtain 50% benchmark students and 50% comprised of

combined strategic and intensive students. First, each grade level list was sorted by instructional recommendation. Second, from the district-wide population, students were randomly sampled from the instructional recommendation groups using the random number function in Excel. At some grade levels, there were not enough students from the strategic and/or intensive groups to meet the sampling goal, and all students from these groups were sent consent forms. To compensate, grade level lists were re-sorted by NWF scores for first grade and DORF scores from second through sixth grade. From these resorted lists, the lowest scoring students were selected to have consent forms sent home. The number of additional consent forms sent home to parents of students in this "low-scoring" group was calculated to bring the total number sent to 46. Those consented students with the lowest scores were selected for testing. Kindergarten was not re-sampled, and 43 consent forms were sent. Fourth grade had enough students in all instructional recommendation groups, and was not re-sampled.

The sampling goal was 30 students per grade (i.e., 15 benchmark students and 15 students from the strategic/intensive combination group). Though the target number of students to test in each group were the same across the strategic and intensive groups, some grade levels may have had more consent forms sent home than reported to ensure the sampling goal was met.

The sample size breakdown by grade level and instructional recommendation based on fall data for students who had consent forms sent home is reported in Table 8. Despite efforts to obtain the desired sample, most of the retesting sample in first, second, third and sixth grades was performing at benchmark according to the fall data.

Table 8

Test-Retest Sample Stratification

Status by grade	N	Consents	Ideal	Actual	Percentage
Kindergarten					
Benchmark	26	26	15	9	53
Strategic	13	13	11	6	35
Intensive	4	4	4	2	12
First Grade					
Benchmark	60	20 (20)	15 (9-15)	22	79
Strategic	6	6	6	6	21
Intensive	0	0	0	0	0
Second Grade					
Benchmark	58	20 (20)	15 (9-15)	19	90
Strategic	4	4	4	2	10
Intensive	2	2	2	0	0
Third Grade					
Benchmark	58	20 (14)	15 (3-15)	20	74
Strategic	9	9	9	5	19
Intensive	3	3	3	2	7
Fourth Grade					
Benchmark	61	20	15	12	57
Strategic	17	17	8	5	24
Intensive	8	8	8	4	19
Fifth Grade					
Benchmark	77	20 (11)	15 (0-15)	13	57

Strategic	9	9	9	6	26
Intensive	6	6	6	4	17
Sixth Grade					
Benchmark	80	20 (16)	15 (5-15)	12	80
Strategic	7	7	7	2	13
Intensive	3	3	3	1	7

Note. District C data only. Stratified sampling goal was 50% of the sample from Benchmark status, and 50% of the sample evenly divided between Strategic and Intensive. N = sample from which student ID numbers were drawn to receive consent forms; Consents = number of students that received consent forms and in parenthesis the number of students from the low-scoring benchmark group who received consent forms; Ideal = the ideal number of students that sent consent forms back and in parenthesis is a range for the number of students needed from the low-scoring benchmark group to bring the sample to 30; Actual = the actual number of students that sent consent forms back (the number of students from the low-scoring group that returned consent forms was unavailable); Percentage = proportion of students that returned consent forms from each group.

GRADE testing. Students were sampled from the fall benchmark data (exported from each site). The desired final sample size for each grade level was 50 students. The sampling goal was to have approximately 50% of the students sampled have benchmark instructional recommendations and 50% be a combined group of students with strategic and intensive instructional recommendations. Attempts were made to accommodate each sites' requests for the manner in which students were to be selected (e.g., by in-tact classrooms).

District A. First, each grade level list was sorted by teacher/classroom. Second, within each teacher's classroom, students were sorted by instructional recommendation. Next, each classroom's composition was reviewed.

District B. In grades K - 5, each grade level list first was sorted by teacher/classroom. Then, within each teacher's classroom, students were sorted by instructional recommendation. Next, each classroom's composition was reviewed. In several grades, there were few students that had strategic or intensive instructional recommendations. For grade 6, students were selected from the entire pool of 6th grade students and not by classroom. First, the list of students with fall DIBELS data was sorted by instructional recommendation. A total of 20 students with strategic instructional recommendations and 7 students with intensive instructional recommendations were in the total 6th grade population. Thus, 37 of the 169 students with Benchmark instructional recommendations were selected randomly using the random number function in Excel.

District C. First, each grade level list was sorted by teacher/classroom. Second, within each teacher's classroom, students were sorted by instructional recommendation. Next, each classroom's composition was reviewed. At several grade levels, few students had strategic or intensive instructional recommendations.

District D. First, each grade level list was sorted by teacher/classroom. Second, within each teacher's classroom, students were sorted by instructional recommendation. Next, each classroom's composition was reviewed. At several grade levels, few students had strategic or intensive instructional recommendations.

District E. Because the participating school from this district was small, the goal was to test all students from each grade level. One student in 2nd grade and one student in 4th grade was removed from the list of eligible students for testing due to developmental disabilities. Each grade-level list was sorted by instructional recommendation. Some students did not complete the academic year at this school, and therefore their data is incomplete across benchmark assessment time points. Some of the students in the school left after fall benchmark testing, and other

students joined the school before or after the winter benchmark assessment. The coordinator at this site reported that the exact number of students from each grade level that left the school was not available. The additional students who entered the school after the fall benchmark testing were sent consent forms in order to maximize variability in the sample. Therefore, the instructional recommendation status of these students for fall is unknown. These numbers are listed in the note of Table 9.

The instructional recommendation status for students was provided by each district. For some students, this information was not provided, and therefore not included in the sampling. The sample size and stratification breakdown by grade level and instructional recommendation based on fall data for students for whom consents were sent home is reported in Table 9. Achieving the sampling goal varied by district. In fall, when examining the entire sample at each grade, for only two grade levels was less than 1/3 of the sample from the lower performing student groups (i.e., strategic and intensive instructional recommendations).

Table 9

Beginning of Year GRADE Testing Sample Stratification by District Instructional Recommendation Status

Status by grade	District A			District B			District C			District D			District E ^a			Totals		
	C	T	%	C	T	%	C	T	%	C	T	%	C	T	%	C	T	%
Kindergarten																		
Benchmark	25	15	56	37	15	56	26	23	58	40	32	74	21	18	43	149	103	58
Strategic	21	6	22	13	9	33	13	13	33	12	6	14	13	11	26	72	45	25
Intensive	18	6	22	5	3	11	18	4	10	8	5	12	14	13	31	63	31	17
First Grade																		
Benchmark	44	25	66	51	41	98	60	45	90	64	40	80	21	17	81	240	168	84
Strategic	14	9	24	5	1	2	6	5	10	12	8	16	3	2	10	40	25	12
Intensive	6	4	11	0	0	0	0	0	0	4	2	4	3	2	10	13	8	4
Second Grade																		
Benchmark	40	24	56	52	43	83	58	29	91	58	37	65	24	22	85	232	155	74
Strategic	21	11	26	8	6	12	4	2	6	11	15	26	13	0	0	57	34	16
Intensive	16	8	19	4	3	6	2	1	3	7	5	9	9	4	15	38	21	10
Third Grade																		
Benchmark	28	10	50	49	32	73	58	33	85	45	35	65	17	16	50	197	126	67

Strategic	17	6	30	9	6	14	9	3	8	20	8	15	9	7	22	64	30	16
Intensive	18	4	20	9	6	14	3	3	8	10	11	20	12	9	28	52	33	17
Fourth Grade																		
Benchmark	23	7	25	35	22	52	47	25	74	53	32	78	24	20	47	182	106	56
Strategic	13	4	14	15	7	17	14	4	12	14	4	10	19	7	16	75	26	14
Intensive	34	17	61	16	13	31	7	5	15	17	5	12	8	16	37	82	56	30
Fifth Grade																		
Benchmark	28	7	37	43	32	65	59	17	65	57	29	62	26	22	63	213	107	61
Strategic	9	4	21	17	11	22	7	6	23	9	10	21	4	4	11	46	35	20
Intensive	29	8	42	7	6	12	6	3	12	12	8	17	11	9	26	65	34	19
Sixth Grade																		
Benchmark	42	3	33	37	3	43	63	16	94	55	20	50	-	-	-	197	42	58
Strategic	16	4	44	30	4	57	6	0	0	14	11	28	-	-	-	66	19	26
Intensive	13	2	22	7	0	0	3	1	6	11	9	23	-	-	-	34	12	16

Note. Stratified sampling goal was 50% of the sample from Benchmark status, and 50% of the sample evenly divided between Strategic and Intensive. C = number of students that received consent forms; T = the actual number of students tested (i.e. the number of students who returned Consent Forms); % = percentage of students in each instructional recommendation category.

^a During the school year, additional students who were not present during fall benchmark testing were sent consent forms in District E:

kindergarten = 1, first grade = 4, second grade = 1, third grade = 3, and fourth grade = 3. The initial status of these students is unknown.

There were additional students tested who were present during fall benchmark assessment but whose initial status was not provided by the school district: 34 students from District B, 28 students from District C, 7 students from District D, and 21 students from District E.

Consumer feedback. At the conclusion of the Benchmark Goals study, teachers and other educators at each school site who (a) administered and scored DIBELS (i.e., assessors); and/or (b) were consumers of the DIBELS data (teachers or administrators) completed the BGS Usability Questionnaire. The questionnaires were completed via the Internet using a third-party web site designed to deliver online questionnaires and surveys. Participants were sent the link (URL) to the questionnaire and completed it at a time convenient for them. Different respondent groups were administered separate items with some overlap across groups. Eleven of the 16 rated items were administered to those in the "assessor only" and "teacher and assessor" respondent groups. One of the 16 items was administered to the "teacher and assessor" respondent group. Four of the 16 rated items were administered to the "teacher only" and "teacher and assessor" respondent groups.

Measures

Measures in this study included the next version of DIBELS measures, the GRADE, and the consumer feedback questionnaire. Each next version of DIBELS measure is described here first, followed by the GRADE, a demographic information form parents of GRADE participants were asked to complete, and the consumer feedback questionnaire. All DIBELS measures are standardized and individually administered, except for Daze, which may be group-administered. The next version of DIBELS measures all follow similar administration procedures and scoring rules used in previous research and practice. The GRADE is a standardized, norm-referenced group-administered achievement test.

DIBELS Measures. We provide brief descriptions of all DIBELS individual measures as well as the stimulus and response demands for each measure in what follows. For more detailed

information on the test construction of individual DIBELS measures see the DIBELS Next Technical Manual.

First Sound Fluency (FSF). First Sound Fluency (FSF) is a brief, direct measure of a student's fluency in identifying the initial sounds in words. The ability to isolate the first sound in a word is an important phonemic awareness skill that is highly related to reading acquisition and reading achievement (Yopp, 1988). The ability to isolate and identify the first phoneme in a word is an easier skill than segmenting all the sounds in words or manipulating phonemes in words. Thus, FSF is used as a measure of developing phonemic awareness at the beginning and middle of kindergarten.

Using standardized directions, the assessor says a series of words one at a time to the student and asks the student to say the first sound in the word. On the scoring page, the assessor circles the corresponding sound or group of sounds the student says. Students receive either 2 points for saying the initial phoneme of a word (e.g., saying the /s/ sound as the first sound in the word street) or 1 point for saying the initial consonant blend, consonant plus vowel, or consonant blend plus vowel (e.g., /st/, /str/, or /strea/ for street). A response is scored as correct as long as the student provides any of the correct responses listed for the word. The total score is the sum of the points the student receives in 1 minute.

To make FSF more sensitive for use with young students, the measure uses differential scoring, which allows students to receive partial credit for demonstrating beginning skills in phonemic awareness. A student who may not be able to isolate the initial phoneme would receive partial credit for providing the first group of sounds in the word, showing emerging understanding that words are made up of sounds. Although partial credit is given, the goal is for

the student to be able to correctly say the first phoneme of each word. This preferred response is given the most points.

Letter Naming Fluency (LNF). Letter Naming Fluency (LNF) is a brief, direct measure of a student's fluency with naming letters. LNF assesses a student's ability to recognize individual letters and say their letter names. Using standardized directions, the assessor presents a page of upper- and lower-case letters arranged in random order and asks the student to name the letters. The assessor marks letter names that are read incorrectly or skipped. The total score is the number of correct letter names that the student says in 1 minute.

The purpose of LNF is to measure students' automaticity with letter naming. Fluency in naming letters is a strong and robust predictor of later reading achievement (Adams, 1990). All letters are included on the LNF materials, but they appear in stratified, random order.

LNF is an indicator of risk rather than an instructional target. While the ability to recognize and name letters in preschool and at the beginning of kindergarten is a strong predictor of later reading achievement (e.g., Badian, 1995; Walsh, Price, & Gillingham, 1988), studies have failed to show that teaching letter names to students enhances their reading ability (e.g., Ehri, 1983) and, in fact, have demonstrated that successful learning of letter-sound correspondences that leads to reading acquisition can occur without knowledge of letter names (Bruck, Genesee, & Caravolas, 1997; Mann & Wimmer, 2002). Because learning letter names is not a powerful instructional target, benchmark goals are not provided for LNF. LNF is a strong predictor, however, so it is included as a part of the DIBELS Composite Score in kindergarten and early first grade.

Phoneme Segmentation Fluency (PSF). Phoneme Segmentation Fluency (PSF) is a brief, direct measure of phonemic awareness. PSF assesses the student's fluency in segmenting a

spoken word into its component parts or sound segments. Using standardized directions, the assessor says a word and asks the student to say the sounds in the word. The assessor underlines each correct sound segment of the word that the student says. A correct sound segment is any different, correct part of the word the student says. The total score is the number of correct sound segments that the student says in 1 minute. For example, if the assessor says the word fish and the student says /f/ /i/ /sh/, the student has completely and correctly segmented the word into all of its component sounds and the score is 3 correct sound segments. If the student says /f/ /ish/, the score is 2 correct sound segments.

Partial credit is given for partial segmentation. A student who is developing phonemic awareness may not yet segment words completely into individual sounds but may segment parts of words. For example, a student who says the first sound of the word sun (/s/) receives 1 point. A student who says the onset and rime (/s/ /un/) receives 2 points and a student who completely and correctly segments all of the individual phonemes in the word (/s/ /u/ /n/) receives 3 points. Note that consonant blends have two or more phonemes that should be produced separately for a student to receive full credit. For example, for the word trap, a student who says /tr/ /a/ /p/ receives partial credit of 3 points, and one who says /t/ /r/ /a/ /p/ receives the full 4 points. Allowing partial credit in scoring increases the sensitivity of the measure, thus making it possible to measure growth from partial to complete segmentation. Although partial credit is given, the preferred response is for students to completely segment words at the phoneme level by the end of kindergarten.

Nonsense Word Fluency (NWF). Nonsense Word Fluency (NWF) is a brief, direct measure of the alphabetic principle and basic phonics. It assesses knowledge of basic letter-sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant (CVC)

and vowel-consonant (VC) words. The test items used for NWF are phonetically regular make-believe (nonsense or pseudo) words. To successfully complete the NWF task, students must rely on their knowledge of letter-sound correspondences and how to blend sounds into whole words. One reason that nonsense word measures are considered to be a good indicator of the alphabetic principle is that “pseudowords have no lexical entry, [and thus] pseudo-word reading provides a relatively pure assessment of students’ ability to apply grapheme-phoneme knowledge in decoding” (Rathvon, 2004, p. 138).

Following a model and a practice item, the student is presented with a sheet of randomly ordered VC and CVC nonsense words (e.g., dif, ik, nop). Standardized directions are used to ask the student to read the make-believe words the best they can, reading either the whole word or saying any sounds they know. For example, if the stimulus word is tof, the student could say /t/ /o/ /f/ or “tof.” The assessor underlines each correct letter sound produced either in isolation or blended together. Whole words read without sounding out are underlined in their entirety.

There are two separate scores reported for NWF:

1. Correct Letter Sounds (CLS) is the number of letter sounds produced correctly in 1 minute. For example, if the student reads dif as /d/ /i/ /f/, the score for Correct Letter Sounds is 3. If the student reads dif as /di/ /f/ or “dif,” the CLS score is also 3.

2. Whole Words Read (WWR) is the number of make-believe words read correctly as a whole word without first being sounded out. For example, if the student reads dif as “dif,” the score is 3 points for CLS and 1 point for WWR. However, if the student reads dif as “/d/ /i/ /f/ dif,” the score is 3 points for CLS but 0 points for WWR.

The goal is for students to read whole words on NWF. However, an advantage of NWF is that the measures provides a means for monitoring the development of the alphabetic principle

and basic phonics as early as the middle of kindergarten, when producing individual letter sounds is the more common response.

DIBELS Oral Reading Fluency (DORF) and Retell. *DIBELS Oral Reading Fluency (DORF)* is a measure of advanced phonics and word attack skills, accurate and fluent reading of connected text, and reading comprehension. The DORF passages and procedures are based on the program of research and development of Curriculum-Based Measurement of reading by Stan Deno and colleagues at the University of Minnesota (Deno, 1989). There are two components to DORF. The first part is oral reading fluency and the second part is passage retell. For the oral reading fluency component, students are given an unfamiliar, grade-level passage of text and asked to read for 1 minute. Errors such as substitutions, omissions, and hesitations for more than 3 seconds are marked while listening to the student read aloud. For benchmark assessment, students are asked to read three different grade-level passages for 1 minute each. DORF is scored by determining the median number of words read correctly and the median number of errors across the three passages. Using the median score from three passages gives the best indicator of student performance over a range of different text and content. The student's accuracy rate is calculated based on the number of words read correctly and the number of errors, using the following formula:

$$Accuracy = 100 \times \frac{\text{median words correct}}{\text{median words correct} + \text{median errors}}$$

The passage retell component follows the reading of each passage, provided that the student has read at least 40 words correct per minute on that passage, or if the assessor believes it is otherwise appropriate. Passage retell is intended to provide a comprehension check for the DORF assessment, and provides an indication that the student is reading for meaning. With a

prompted passage retell, it is clear to the student that the intent is to read for meaning. Speed-reading without attending to text comprehension is undesirable and will be less likely when the student is told that they will be asked to tell about what they will read.

Case studies have documented students who can read words but not comprehend what they read (Dewitz & Dewitz, 2003). Students who display similar reading behavior may not be identified without a comprehension check. Passage retell provides both a valuable indicator of reading comprehension as well as an efficient procedure to identify those students who have difficulty talking about what they have read. Importantly, the inclusion of the passage retell component explicitly instructs students to be reading fluently *for meaning*. The quality of a student's retell provides valuable information about overall reading proficiency and oral language skills.

During retell, the student is asked to tell about what he/she has read. The assessor indicates the number of words in the retell that are related to the story by drawing a line through a box of numbers. Following a hesitation of 3 seconds, students are prompted to tell as much as they can about the story. If the student hesitates again for 5 seconds or longer, or if the student is clearly responding for 5 seconds in a way that is not relevant to the passage, the task is discontinued. The assessor must make a judgment about the relevance of the retell to the story while drawing the line. A quality of response rating allows the assessor to make a qualitative rating of the student's response after the completion of the retell. The rating should be based on how well the student retold the portion of the passage that he/she read.

DIBELS-Maze (Daze). Daze is the standardized, *DIBELS* version of a maze testing procedure for measuring reading comprehension. The purpose of a maze assessment is to measure the reasoning processes that underlie comprehension. Specifically, Daze assesses the

student's ability to construct meaning from text, word recognition skills, background information and prior knowledge, familiarity with linguistic properties such as syntax and morphology, and reasoning skills. Daze can be given to a whole class at the same time, to a small group of students, or individually. Students are administered a passage in which approximately every seventh word has been replaced by a box containing the correct word and two distractor words. Using standardized directions, students are asked to read the passage silently and circle their word choices. The student receives credit for selecting the word that best fits the passage from the choices available in each box presented in the reading passage. The scores that are recorded are the number of correct and incorrect responses. An adjusted score, which compensates for guessing, is calculated based on the number of correct and incorrect responses. The following procedure is used to calculate the adjusted score:

$$\text{Daze Adjusted Score} = \text{number of correct responses} - (\text{number of incorrect responses} \div 2).$$

The result of this calculation is rounded to the nearest whole number, with half-points (0.5) rounded up. If the adjusted score is negative, a score of 0 is recorded. Since there are three possible responses for each item, the adjusted score expected for a student who guesses on every item is 0. Scores are not pro-rated. The maximum score a student can receive is equal to the number of items on the form.

DIBELS Composite Score (DCS). The DIBELS Composite Score is a combination of multiple DIBELS scores. In creating the DIBELS Composite Score, we used an improper linear model to equally weight the *DIBELS Next* measures that are strongly correlated to important outcomes. Robyn Dawes (1979) describes the robust beauty of improper linear models (ILM) to make important social decisions. With proper linear models, multiple variables are combined to provide the “best” or optimal prediction of a target outcome of interest. For example, in multiple

regression, coefficients are optimal in that they maximize the amount of explained variance in the outcome variable. Such models are better than reliance on clinical judgment. A disadvantage to proper linear models is that, although they provide the best prediction for a specific sample of subjects on a specific target outcome variable, they may not provide the best prediction for a generalization sample of subjects or for a generalization outcome. Here we use target outcome to refer to the specific outcome variable used to derive the linear regression coefficients and the selected sample refers to the specific sample of subjects used to compute the coefficients. A generalization outcome refers to a different measure of the same or closely related construct, and a generalization sample refers to a different sample from the same or similar population of interest.

Furthermore, while multiple regression analysis provides the best prediction of the target outcome in the sense of most variance explained, there are other ways that “best” could be operationalized. For example, a different set of coefficients may best discriminate students who are making adequate progress from students who are not (linear discriminant function analysis). Another set of coefficients would best capture the variance represented by multiple variables (principle components analysis). Still another different set of coefficients would best represent the shared variance of multiple variables (factor analysis).

Dawes (1979) proposes a radical and elegant alternative to a proper linear model: an improper linear model (ILM) using non-optimal coefficients. According to Dawes, three potential non-optimal methods to consider are intuition, simulation, or unit weighting (i.e., weights are set to be equal). Unit weighting (equal weighting) has the advantage of being quite robust and generalizes well across samples, outcomes, and operationalizations of “best.” A unit weighted, improper linear model is constructed by first selecting the variables that have

substantial, positive correlations with the target outcome. Then the scores are standardized to have mean of 0 and standard deviation of 1. Finally the standardized scores are summed or averaged to form a composite score.

According to Dawes, an advantage of a unit-weighted ILM is that the resulting composite is more likely to generalize to:

- different outcome variables representing the same or similar constructs,
- different samples of subjects from the same or similar populations, and
- different ways of operationalizing "best."

Especially when predicting to an outcome for which there is a lack of agreement about a single best criterion or outcome, a unit-weighted ILM may even be a better choice than a proper linear model. According to Dana (2010) ILM may also be the better option compared to a PLM when predicting beyond a specific sample. Because the amount of difference between the sample studied and the population can be a significant source of error, an improper linear model can provide a better prediction to new and different samples.

For each grade and time of year, the DIBELS Next measures that correlate highly with later outcomes are combined into a DIBELS Composite Score (DCS). Each measure is weighted so that all contribute approximately equally to the DCS. Weighted scores have approximately equal standard deviations. The DCS represents a rich and broad sample of behavior and provides a more complete sample of reading behavior than any single measure. For example in third grade beginning of year, DORF Words Correct, DORF Accuracy, DORF Retell, and Daze all correlate strongly and positively to important reading outcomes. A unit-weighted improper linear model score for student i would be computed using:

$$ILM_i = \frac{X_{1i} - \bar{X}_1}{s_1} + \frac{X_{2i} - \bar{X}_2}{s_2} + \frac{X_{3i} - \bar{X}_3}{s_3} + \frac{X_{4i} - \bar{X}_4}{s_4}$$

Where X_1 = DORF Words Correct, X_2 = DORF Accuracy, X_3 = DORF Retell, and X_4 = Daze.

The terms of the equation can be rearranged so that:

$$s_1 * ILM + constant = X_{1i} + \frac{s_1}{s_2} X_{2i} + \frac{s_1}{s_3} X_{3i} + \frac{s_1}{s_4} X_{4i}$$

Using estimates of the standard deviation from the benchmark goals study for third grade beginning of year, the equation would be:

$$s_1 * ILM + constant = X_{1i} + 4.46X_{2i} + 2.02X_{3i} + 5.76X_{4i}$$

For DORF Accuracy we used a look-up table to obtain appropriately-scaled values, and we used integer weights for the other measures to obtain the DIBELS Composite Score (DCS) such that the DCS would closely approximate a linear transformation of the Dawes unit-weighted improper linear model:

$$DCS_i = X_{1i} + \text{Accuracy\_value}_i + 2X_{3i} + 4X_{4i}$$

A weight of 4 was selected for Daze as most representative of all of the different grade levels and times of year. Calculation of the DIBELS Composite Score, can be done manually with the DIBELS Next Composite Score Worksheets (see Appendix B) or using a recommended data reporting service.

As illustrated in Table 10 the DIBELS Composite Score explains more variance in reading outcomes than does DORF Words Correct alone. Across grades 1 through 6, the median *additional* variance explained is 9%, ranging from 3% to 17%, generally with greater additional variance explained in the upper grades. In other words, although DORF Words Correct alone is very good, the DIBELS Composite Score is even better in meaningful and important ways.

Table 10

Correlation of DIBELS Composite Score Compared to DORF Words Correct in Predicting End of Year GRADE Total Test Score with Additional Variance Explained

Grade and Time of Year	DORF Words Correct Predicting GRADE Total	DIBELS Composite Score Predicting GRADE Total	Additional Variance Explained by DIBELS Composite Score
Grade 1 Middle of Year	.64	.70	8%
Grade 1 End of Year	.75	.77	4%
Grade 2 Beginning of Year	.69	.75	8%
Grade 2 Middle of Year	.76	.80	5%
Grade 2 End of Year	.73	.75	3%
Grade 3 Beginning of Year	.66	.73	10%
Grade 3 Middle of Year	.67	.78	15%
Grade 3 End of Year	.66	.75	13%
Grade 4 Beginning of Year	.76	.80	5%
Grade 4 Middle of Year	.76	.80	6%
Grade 4 End of Year	.75	.80	8%
Grade 5 Beginning of Year	.69	.76	11%
Grade 5 Middle of Year	.64	.76	17%
Grade 5 End of Year	.66	.77	17%
Grade 6 Beginning of Year	.64	.71	9%
Grade 6 Middle of Year	.59	.68	12%
Grade 6 End of Year	.61	.73	16%

Note. DORF = DIBELS Oral Reading Fluency.

In addition to providing the strongest and most generalizable prediction of reading outcomes, the DCS also conveys that all of the aspects of reading proficiency are critical. A student whose DCS is At or Above Benchmark is reading at an adequate rate, with a high degree of accuracy, and attending to the meaning of the passage. Students who score at or above the DCS benchmark goal may still need additional support in one or more of the basic early literacy skills. However, the DCS provides the best overall estimate of the student's reading proficiency. Thus, the DCS is the first score to be evaluated when determining the overall likelihood of a student meeting future reading goals. In this study, we used the DCS as an internal criterion measure for developing benchmark goals for individual DIBELS measures.

Group Reading Assessment and Diagnostic Evaluation (GRADE)

The GRADE is an un-timed group-administered norm-referenced reading achievement test appropriate for children in preschool up through grade 12. The GRADE is comprised of 5 components. These 5 components are comprised of 16 subtests. Not all 16 subtests are used at each testing level. Various subtest scores are combined to form the following composite scores: Phonemic Awareness, Early Literacy Skills, Comprehension, Vocabulary, and Total Test. There are 11 testing levels in the GRADE, and most correspond directly to specific grade levels (i.e., grades K - 6). Brief descriptions of each subtest and the grade levels at which they are given are summarized in Table 11.

For this study, students in kindergarten were administered 7 subtests over the course of 3 test sessions for a total testing time of approximately 90 minutes. Students in first and second grade were administered 5 subtests across two test sessions for a total testing time of about 85 minutes.

Students in third grade were administered 5 subtests in two sessions for a total testing time of about 65 minutes. Students in fourth through sixth grade were administered 3 subtests across two testing sessions for a total testing time of about 60 minutes at each of these grade levels.

Overall, the GRADE has excellent reliability and validity for its intended purposes. The GRADE technical manual (Williams, 2001) indicates that coefficient alpha and split-half reliabilities for the Total Test across grades preschool through grade 12 range from .89 -.96 and .94 -.98, respectively (Table 12). Alternate form reliabilities range from .81 -.94 across preschool to grade 12 for the Total Test, while test-retest reliabilities range from .77 -.96 (Williams, 2001). Correlations of the Total Test with other group administered achievement tests (e.g., ITBS Total Reading Standard Score) range from .69 -.90. The correlations with the Peabody Individual Achievement Test-Revised Total Reading portion range from .74 -.80. Predictive validity data indicate that the GRADE Total Test score in the fall was correlated .76, .77, and .86 with TerraNova scores in the spring for grades 2, 4, and 6, respectively.

Table 11

Subtest Descriptions and Grade Levels in Benchmark Goals Study

GRADE Subtest	Description	Grade	Composites Contributed To
Sound Matching	The student is required to select from an array of pictures the one picture that corresponds to the beginning or ending sound of a word presented orally by the examiner.	Kindergarten	Phonological Awareness
Rhyming	The examiner orally presents a target word and then names four pictures. The student is required to mark the picture that rhymes with the target word.	Kindergarten	Phonological Awareness
Print Awareness	The student is required to select the picture from an array that best fits a sentence presented by the examiner. Student is required to discriminate letters, numbers, words, & sentences.	Kindergarten	Early Literacy Skills
Letter Recognition	Examiner says a letter name and the student is required to choose the correct letter from an array of five letters.	Kindergarten	Early Literacy Skills
Same & Different Words	Examiner points to a word in a box and the student is required to find the word that is the same (or different) in an array of four words.	Kindergarten	Early Literacy Skills
Phoneme-Grapheme Correspondence	Examiner says a word and the student is required to mark the letter in an array of 4 letters that makes the either the first or last sound in the word the examiner said.	Kindergarten	Total Test Composite
Word Reading	Examiner reads a target word, reads a sentence that contains the word, then repeats the word. The student is required to mark the target word within an array of four or five words.	First Second Third	Vocabulary Composite Total Test Composite

Word Meaning	The student is required to silently read a target word and then select a picture from an array of 4 pictures that best represents the meaning of the target word.	First Second	Vocabulary Composite Total Test Composite
Vocabulary	The student is required to read a sentence or phrase in which one word is in bold print and then choose the meaning of that bolded word from a list or four or five choices.	Third Fourth Fifth Sixth	Vocabulary Composite (Grade 3 only) Total Test Composite
Sentence Comprehension	The student is required to read a sentence in which a word is replaced by a blank, then choose the correct word to fill in the blank from an list of four or five choices.	First Second Third Fourth Fifth Sixth	Comprehension Composite Total Test Composite
Passage Comprehension	The student is required to read a passage of one or more paragraphs and then answer three to five multiple choice questions about the passage.	First Second Third Fourth Fifth Sixth	Comprehension Composite Total Test Composite
Listening Comprehension	The examiner reads a sentence aloud and the student is required to mark one of four pictures the goes best with what the examiner read.	Kindergarten First Second Third	Total Test Composite for grade K only.

Note. GRADE Technical Manual, Williams, K.T., Ph.D. (2001).

Table 12

Group Reading Assessment and Diagnostic Evaluation (GRADE) Criterion Measures Reliability

Composite and Subtest(s)	Cronbach's Alpha and Split-Half Reliability by grade													
	K		1		2		3		4		5		6	
	A	S	A	S	A	S	A	S	A	S	A	S	A	S
Total Test	.89	.94	.96	.98	.96	.98	.95	.97	.94	.97	.95	.97	.95	.97
Phonological Awareness	.87	.93	-	-	-	-	-	-	-	-	-	-	-	-
Early Literacy Skills	.66	.80	-	-	-	-	-	-	-	-	-	-	-	-
Phoneme-Grapheme Corresp.	.80	.89	-	-	-	-	-	-	-	-	-	-	-	-
Comprehension Composite	-	-	.93	.96	.96	.98	.89	.94	.91	.95	.93	.96	.93	.96
Sentence Comprehension	-	-	.90	.95	.94	.97	.83	.91	.86	.92	.89	.94	.88	.94
Passage Comprehension	-	-	.88	.94	.87	.93	.83	.91	.85	.92	.88	.94	.90	.95
Vocabulary Composite	-	-	.92	.96	.93	.96	.91	.95	-	-	-	-	-	-
Word Reading ^a	.73	.84	.81	.90	.88	.94	.82	.90	-	-	-	-	-	-
Word Meaning	-	-	.90	.95	.88	.94	-	-	-	-	-	-	-	-
Vocabulary	-	-	-	-	-	-	.89	.94	.85	.92	.84	.91	.85	.92
Listening Comprehension ^b	.71	.83	.76	.86	.60	.75	.52	.68	-	-	-	-	-	-

Note. A = Cronbach's Coefficient Alpha. S = Split-Half Reliability corrected by the Spearman-Brown formula for full-test length. K = Kindergarten, 1 - 6 = first through sixth grades. Phonemic Awareness is made up of Sound Matching and Rhyming subtests. Early Literacy Skills is made up of Print Awareness, Letter Recognition and Same & Different Words. Comprehension Composite is made up of Sentence and passage Comprehension. Vocabulary Composite is made up of Word Meaning and Word Reading (grades 1 and 2), Word Reading and Vocabulary (grade 3). Total Test Composite is made up of the subtests that contribute to the standard scores for their respective grade levels. Vocabulary exists as a subtest to the Vocabulary Composite for grade 3, and as a subtest to Total Test Composite for grades 4, 5, and 6. *GRADE Technical Manual*, Williams, K.T., Ph.D. (2001).

^a For kindergarten, Word Reading is optional and not included in Total Test Composite at level indicated.

^b Not included in Total Test Composite at level indicated.

Table 13. *The Relationship of the GRADE Criterion Measures to DIBELS Measures*

GRADE Measures	DIBELS Measures							
	LNF	FSF	PSF	CLS	WWR	DORF	Retell	Daze
Total Test	X	X	X	X	X	X	X	X
Phonological Awareness		X	X					
Early Literacy Skills	X			X	X			
Phoneme-Grapheme Corresp.				X	X			
Comprehension Composite				X	X	X	X	X
Sentence Comprehension				X	X	X	X	X
Passage Comprehension				X	X	X	X	X
Vocabulary Composite								
Word Reading						X		X
Word Meaning								
Vocabulary								
Listening Comprehension ^a								

Note. X = The GRADE measure was evaluated for criterion-related validity for the DIBELS measure. LNF = Letter Naming Fluency; FSF = First Sound Fluency; PSF = Phoneme Segmentation Fluency; CLS = Nonsense Word Fluency Correct Letter Sounds; WWR = Nonsense Word Fluency Whole Words Read; DORF = Oral Reading Fluency; Daze = Daze Adjusted Score.

^a Not included in Total Test Composite at level indicated.

DIBELS BGS Usability Questionnaire. This questionnaire was developed in-house and is similar to other questionnaires used to evaluate DIBELS measures and products. The questionnaire includes statements like, “The administration and scoring rules for the next version of DIBELS are easy to follow,” “The next version of DIBELS materials are organized for efficient administration,” “Overall, the data obtained from the next version of DIBELS are helpful for goal setting,” “Overall, the data obtained from the next version of DIBELS are helpful for identifying instructional targets,” and “I would suggest use of the next version of DIBELS to others.” For 16 of the items, respondents were asked to rate statements on a six point Likert-type scale ranging from strongly disagree (1) to strongly agree (6). The questionnaire included one yes/no response question, directed to teachers only. Several open-ended response items provided opportunities for respondents to elaborate on rated statements and comment about specific DIBELS measures (i.e., “please comment on the administration and scoring rules for individual measures”).

Research Questions

The DIBELS Next Benchmark Goals Study was focused on answering four specific questions:

1. What level of performance on DIBELS Next assessments predict a student is likely to achieve future reading goals and what level of performance on DIBELS Next assessments predict a student is unlikely to meet future reading goals?
2. What is the reliability (test-retest, alternate form, and inter-rater) of the DIBELS Next assessments and the DIBELS Composite Score?

3. What are the correlations for DIBELS Next assessments with each other and with the Group Reading Assessment and Diagnostic Evaluation (GRADE), a criterion measure of reading proficiency that includes comprehension?
4. To what extent are teachers and assessors satisfied with the assessments?

Results

This section presents the results of analyses of the DIBELS Next measures in relation to each other, the DIBELS Composite Score, and the external criterion, the GRADE. Extensive data management procedures are first described, followed by sample size descriptions for the various subsets of data. Second, descriptive statistics and correlations for individual DIBELS Next measures and the DIBELS Composite Score are presented. Third, data regarding the reliability and validity of DIBELS Next are presented. Following the validity data, we present data on benchmark goals and cut-points for risk for the DIBELS Next measures and the DIBELS Composite Score. Finally, data from the DIBELS BGS Usability Questionnaire are presented.

Data Management Procedures

Due to the large sample size, number of school districts, and various data sets involved in the analyses for the BGS, detailed data management procedures were followed to assure that we were working with a corrected and clean data set. Data was imported from the research database after each benchmark period or reliability testing period after notification from the districts that all their data had been entered. Data were then checked at the item level for complete records, including scores and student ID numbers. Any record without a student ID or without scores was removed and the district was notified. Additionally, all IDs with duplicates were removed, resulting in a final sample of 3,816 students with data on at least one DIBELS measure for one or more time points. We also removed data for scores that were invalid due to known errors with

data collection (e.g., lack of fidelity to assessment procedures), invalid ranges (i.e., scores above the maximum possible for a given measure), or significant univariate or bivariate outliers. See measure-specific sections for details regarding the way in which these steps impacted our final sample sizes by measure, grade, and time of year.

Scoring Decision Rules and Outliers. We examined all of the DIBELS Next data by measure, grade, and time of year for the presence of outliers and/or invalid scores. Almost every measure by grade and time of year had scores that were more than 1.5 standard deviations above the mean, as is common in large datasets (Tabachnick & Fidell, 2007). We therefore defined outlier scores more stringently as scores that were *more than three standard deviations above the mean score* for that measure. What follows is a detailed description of the decision rules for deleting outliers from each grade by time of year by measure. Removal of outliers in individual measures decreased the sample size for the analyses including that measure, but did not affect the overall sample size of the study.

Nonsense Word Fluency. Outlier scores for NWF were examined from two perspectives. First, we considered the CLS score relative to total words attempted (TW). Second, CLS and WWR were examined together. Because nonsense words are two or three letters in length, the number of total words attempted should be greater than or equal to a third of the correct letter sounds attempted. Examination of the data revealed 30 invalid pairs of scores for CLS and TW in the middle of kindergarten, 13 in the end of year kindergarten, 13 in the beginning of first grade, 9 in the middle of first grade, 4 in the end of year of first grade, and 7 in the beginning of second grade. Total words was a variable new to this study, and there may have been some confusion as to what was to be counted as a total word. Therefore, where there was an invalid pair of scores, only the total words attempted (TW) score was removed from the data set.

If a student read a word as a whole word (WWR), then the letters in that whole word read were counted toward the CLS score (i.e., two to three CLS per whole word read). Therefore, if a student had a score other than 0 for WWR, their CLS score should be at least twice their WWR score (at minimum, a student should receive 2 CLS points per word read whole). Under these conditions, their CLS score should also be less than three times their WWR score (at maximum, a student should receive 3 CLS points per whole word read). Three WWR scores from middle of year kindergarten, 1 from the middle of year in first grade, and 2 from the end of year in first grade were deemed invalid and removed from the data set because they did not fall within these parameters.

Daze. Two scores contribute to the Daze score: a score that indicates the number of words correctly chosen and an errors score that indicates were chosen incorrectly to complete the sentence. The sum of these two scores should not be greater than the number of items in the measure. Any pair of scores on this measure that violated this rule were removed from the data set. Using this decision rule, we removed 1 pair of scores from the middle of year in fourth grade, 1 pair at the end of year in fourth grade, 1 pair at the beginning of year in fifth grade, and 1 pair at the end of the year in fifth grade.

Daze Adjusted Score. The number of Daze correct responses was adjusted to more accurately reflect the students' overall performance by factoring in the number of incorrect responses. This adjustment was completed by subtracting half the number of incorrect responses from the number of correct responses. In some cases, this adjustment resulted in a negative Daze Adjusted Score. A negative score is the result of the student getting twice as many items incorrect on the measure than correct. These scores were changed to zero in the final data set. A total of 28

negative Daze Adjusted Scores were set to zero. A more detailed account of the negative Daze Adjusted Scores by time of year is reported next.

Fall BGS data had twelve negative Daze adjusted scores ranging from -1 to -6. These scores were distributed as follows: nine in third grade, one in fourth grade, and two in fifth grade. No negative Daze scores were found in sixth grade. Of those students with negative Daze adjusted scores, the mean raw Daze score was 3 in third grade, 1 in fourth grade, and 4 in fifth grade; average Daze errors made was 8 in third grade, 5 in fourth grade, and 17 in fifth grade. An examination of these student's DORF scores revealed all students with negative Daze adjusted scores were below benchmark on DORF. Mean DORF scores were 27 for third grade, 13 for fourth grade, and 36 for fifth grade. A total of 12 Daze adjusted scores from fall BGS data were set to zero.

Winter BGS data had six negative Daze adjusted scores ranging from -1 to -6. These scores were distributed as follows: one in third grade, two in fourth grade, one in fifth grade, and two in sixth grade. For students with negative Daze scores, the mean raw Daze score was 0 in third grade, 6.5 in fourth grade, 3 in fifth grade, and 2.5 in sixth grade; mean Daze errors made was 6 in third grade, 21.5 in fourth grade, 11 in fifth grade, and 7 in sixth grade. Mean DORF scores for these students were 12 for third grade, 22.5 for fourth grade, 42 for fifth grade, and 41.5 in sixth grade. A total of 6 Daze adjusted scores from winter BGS data were set to zero.

Spring BGS data had ten negative Daze adjusted scores ranging from -1 to -8. Five of these negative scores were found in third grade, one in fourth grade, three in fifth grade, and one in sixth grade. For students with negative Daze scores, the mean raw Daze score was 7.8 in third grade, 7 in fourth grade, 3 in fifth grade, and 2.5 in sixth grade; mean Daze errors made was 6 in third grade, 21.5 in fourth grade, 13.3 in fifth grade, and 2 in sixth grade. Mean DORF scores for

these students were 49.6 for third grade, 27 for fourth grade, 77.6 for fifth grade, and 20 in sixth grade. A total of 10 Daze adjusted scores from spring BGS data were set to zero.

Inter-rater, test-retest, alternate-form reliability data. Additional reliability testing using DIBELS measures was performed during all three times of year. The same requirements for scoring as the benchmark testing were used as well as procedures for determining any invalid scores. Few invalid scores were found. However, a single pair of invalid Daze and Daze errors scores was found in the Inter-Rater Reliability data set in fifth grade. These scores exceeded the maximum possible score on the measure and were therefore removed from the data set. In addition, one invalid Retell Quality of Response rating was found in the Test-Retest Reliability data set in fourth grade. This score was removed from the data set.

Finally, the Alternate-Form data set contained five invalid scores. All of the invalid scores were from the Retell Quality of Response portion of DORF. One of these invalid scores was in first grade and one was in second grade. The other three scores were entered with the reliability data for middle of year Retell Quality of Response. All five invalid scores were removed from the data set.

GRADE. Examination of the GRADE data set revealed 63 scores across multiple grades and sub-tests that were classified as outliers. All scores were valid and did not exceed the maximum possible value and therefore remained in the data set for further analyses.

Descriptive Statistics

The descriptive statistics (mean and standard deviation) for each DIBELS measure by grade and time of year in Table 15. Scores for most measures given at more than one point during the year increase over the course of the year. The standard deviation data for many measures

suggests increased variability in scores over the course of the year at several grade levels, but in particular in the upper grades.

Quartiles for each DIBELS measure by grade and time of year are provided in Table 18. When comparing the mean score with the corresponding median score for each measure, we noted a negative skew in the distribution of scores for the following: kindergarten middle of year PSF, first grade end of year DORF accuracy, second grade beginning of year ORF accuracy, fourth grade middle of year ORF, fifth grade beginning and end of year ORF, and sixth grade beginning of year ORF. We also noted evidence of positive skew in the distribution of scores on NWF: CLS and WWR at the end of year in kindergarten through the beginning of year in second grade.

Descriptive statistics for the DIBELS Composite score, including quartiles, are presented in Table 18. We noted steady growth in the mean composite score in all grades, and across time points.

Detailed percentages of the population at each benchmark and cut point for risk are presented in Table 14 for each individual measure and for the DIBELS Composite Score. The percent of students at or above benchmark on the composite score ranged from 64% to 79%. In practice, the typical percentage of students at or above benchmark is around 60. These data suggest that our sample was above average overall based on the percent above benchmark on the DIBELS Composite Score for each grade by time of year combination (see Table 14). The percent of students at benchmark on individual measures ranged from 45% for DORF Accuracy at the end of the year in fifth grade, to 100% in the middle of the year of third grade for Retell, with an average of 71% of students scoring at or above benchmark on a single measure. Figure 2, page

157, provides the updated benchmark goals and cut points for risk for DIBELS measures and the DIBELS Composite Score.

Because DIBELS were designed for both universal screening and progress monitoring, we evaluated the rate of change/growth on the measures throughout the course of the school year. The growth for each measure by grade and time of year is illustrated in Table 17, which displays descriptive statistics on the change in scores across times of the year. The growth over the entire school year appears to be nonlinear, with a larger increase in scores between the beginning and middle of the school year than between the middle and end of the year. The DORF errors statistics are interpreted differently from the others presented in the table. When examining DORF errors, positive growth is noted by the decrease in this number. A decrease in errors indicates increased accuracy, which is a desirable or positive outcome.

Table 14

BGS Percent of Students at Benchmark (Percent Signs Omitted)

Measure by Grade	Beginning of Year			Middle of Year			End of Year		
	Well Below	Below	Benchmark	Well Below	Below	Benchmark	Well Below	Below	Benchmark
<i>Kindergarten</i>									
FSF	20	9	71	7	15	78	-	-	-
PSF	-	-	-	9	15	76	4	14	81
NWF-CLS	-	-	-	12	22	67	9	29	62
DIBELS Composite Score	17	16	67	11	21	68	9	23	68
<i>First Grade</i>									
PSF	12	23	65	-	-	-	-	-	-
NWF-CLS	11	22	66	18	18	64	18	15	67
NWF-WWR	-	32	68	18	16	66	14	17	69
DORF Words Correct	-	-	-	15	17	68	17	14	69
DORF Accuracy	-	-	-	18	17	65	15	15	70
Retell	-	-	-	-	-	-	-	19	81
DIBELS Composite Score	18	13	68	18	15	67	18	15	68

<i>Second Grade</i>									
NWF-CLS	15	23	62	-	-	-	-	-	-
NWF-WWR	22	19	59	-	-	-	-	-	-
DORF Words Correct	21	12	67	19	13	68	13	18	68
DORF Accuracy	14	16	70	21	0	79	16	0	84
Retell	8	13	79	9	14	77	7	17	77
DIBELS Composite Score	17	11	72	17	10	73	13	14	73
<i>Third Grade</i>									
DORF Words Correct	16	12	71	1	0	99	1	0	99
DORF Accuracy	13	21	66	9	18	72	15	16	69
Retell	6	19	75	0	0	100	0	0	100
Daze Adjusted Score	15	14	71	14	16	69	14	17	69
DIBELS Composite Score	18	10	72	16	11	73	14	13	72
<i>Fourth Grade</i>									
DORF Words Correct	21	14	65	15	17	68	15	18	67
DORF Accuracy	20	21	59	18	0	82	10	26	64
Retell	11	21	68	14	23	64	13	17	69
Daze Adjusted Score	16	20	64	16	27	57	14	18	68

DIBELS Composite Score	21	12	67	21	12	68	14	18	68

Fifth Grade

DORF Words Correct	22	14	64	15	22	63	16	21	64

DORF Accuracy	17	28	55	17	0	83	17	38	45

Retell	14	24	62	12	23	65	12	24	63

Daze Adjusted Score	15	20	65	14	32	53	14	21	65

DIBELS Composite Score	13	22	65	14	21	65	14	22	64

Sixth Grade

DORF Words Correct	10	12	78	12	10	79	99	13	78

DORF Accuracy	10	16	75	10	0	90	11	20	70

Retell	4	17	80	4	17	79	9	11	80

Daze Adjusted Score	10	12	78	8	12	80	8	17	76

DIBELS Composite Score	9	13	78	9	12	79	9	12	79

Note. Sample size = 3,816.

Table 15

BGS Descriptive Statistics for DIBELS Next Measures

Beginning of year				Middle of year			End of year		
Measure	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>
Kindergarten									
FSF	18.10	12.75	478	38.19	12.08	476	-	-	-
LNF	22.35	15.72	478	41.92	16.34	476	50.54	15.64	479
PSF	-	-	-	37.71	18.55	476	50.52	14.70	479
NWF-CLS	-	-	-	23.69	14.56	476	36.77	21.27	477
NWF-WWR	-	-	-	1.93	4.08	474	6.02	8.42	477
First Grade									
LNF	49.56	15.97	461	-	-	-	-	-	-
PSF	43.06	15.23	461	-	-	-	-	-	-
NWF-CLS	38.82	25.31	461	59.86	32.52	458	77.08	33.86	461
NWF-WWR	6.93	9.81	461	15.52	13.50	457	21.63	14.50	461
DORF	-	-	-	45.08	35.14	458	66.73	35.00	459

DORFA	-	-	-	.81	.16	451	.90	.12	459
Retell	-	-	-	12.14	15.14	271	27.39	15.09	373
Second Grade									
NWF-CLS	68.53	33.44	560	-	-	-	-	-	-
NWF-WWR	18.22	13.93	560	-	-	-	-	-	-
DORF	70.44	37.20	560	88.38	37.69	565	103.21	38.96	566
DORFA	.91	.12	558	.95	.09	564	.96	.07	566
Retell	26.88	15.95	520	31.26	17.35	548	38.61	18.48	541
Third Grade									
DORF	89.56	37.64	502	103.55	37.44	504	116.84	37.74	498
DORFA	.94	.09	502	.96	.06	502	.97	.06	498
Retell	30.98	16.86	479	38.00	19.77	492	42.98	19.01	486
Daze	11.40	6.58	501	15.39	8.85	503	22.02	8.54	497
Fourth Grade									
DORF	101.15	37.71	589	114.91	38.16	596	129.62	37.41	589
DORFA	.95	.07	589	.96	.06	592	.97	.05	589
Retell	35.61	18.94	577	37.91	19.40	592	44.05	21.04	582
Daze	16.91	7.49	585	18.64	7.85	589	26.65	8.80	586

Fifth Grade									
DORF	121.47	38.32	546	132.02	36.09	543	139.50	38.71	539
DORFA	.97	.06	546	.98	.04	542	.98	.03	539
Retell	38.99	17.53	540	43.53	19.01	541	45.15	19.73	530
Daze	20.39	8.83	543	20.01	8.03	537	27.11	10.06	536
Sixth Grade									
DORF	131.33	35.26	535	136.03	38.10	528	143.93	37.16	531
DORFA	.97	.04	535	.97	.04	528	.98	.03	531
Retell	41.52	17.58	529	47.80	21.05	526	51.63	21.33	527
Daze	22.41	7.65	535	26.52	9.89	520	27.00	9.89	529

Note. FSF = First Sound Fluency; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; NWF-CLS = Nonsense Word Fluency - Correct Letter Sounds; NWF-WWR = Nonsense Word Fluency - Whole Words Read; DORF = DIBELS Oral Reading Fluency; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score.

Table 16

BGS Quartiles for DIBELS Next Measures

Measure	Beginning of year						Middle of year						End of year					
	Min	25th	50th	75th	Max	<i>n</i>	Min	25th	50th	75th	Max	<i>n</i>	Min	25th	50th	75th	Max	<i>n</i>
Kindergarten																		
FSF	0	7	18	28	56	477	0	31	39	46	60	476	-	-	-	-	-	-
LNF	0	9	22	35	70	477	0	30	41	54	91	476	8	40	51	61	100	479
PSF	-	-	-	-	-	-	0	22	41	52	74	476	0	43	51	62	80	479
CLS	-	-	-	-	-	-	0	14	22	32	107	476	0	22	32	47	140	477
WWR	-	-	-	-	-	-	0	0	0	2	33	474	0	0	2	10	42	477
First Grade																		
LNF	0	40	51	60	99	461	-	-	-	-	-	-	-	-	-	-	-	-
PSF	0	36	45	53	77	461	-	-	-	-	-	-	-	-	-	-	-	-
CLS	0	23	33	47	143	461	0	38	53	76	143	458	8	51	70	98	143	461
WWR	0	0	2	10	50	461	0	4	12	23	50	457	0	10	19	31	50	461
DORF	-	-	-	-	-	-	0	19	33	64	211	458	5	39	63	91	195	459
DORFE	-	-	-	-	-	-	0	3	6	8	24	457	0	2	4	6	28	459

DORFA	-	-	-	-	-	-	.00	.72	.84	.96	1.00	451	.39	.89	.95	.98	1.00	459
Retell	-	-	-	-	-	-	0	0	6	22	84	271	0	17	25	34	94	373

Second Grade

CLS	0	42	64	90	143	560	-	-	-	-	-	-	-	-	-	-	-	-
WWR	0	6	16	28	50	560	-	-	-	-	-	-	-	-	-	-	-	-
DORF	0	40	67	93	196	560	0	60	89	117	208	565	3	78	103	129	228	566
DORFE	0	2	4	6	25	559	0	1	2	4	29	565	0	1	2	4	17	566
DORFA	.23	.88	.95	.98	1.00	558	.29	.93	.98	.99	1.00	564	.36	.96	.98	.99	1.00	566
Retell	0	16	26	35	94	520	0	19	28	41	94	548	4	26	35	50	110	541

Third Grade

DORF	4	61	91	112	221	502	0	81	103	129	211	504	11	93	117	145	224	498
DORFE	0	2	3	5	19	502	0	1	2	4	12	504	0	1	2	4	20	498
DORFA	.33	.93	.96	.98	1.00	502	.47	.95	.98	.99	1.00	502	.50	.96	.98	.99	1.00	498
Retell	0	19	28	41	94	479	0	25	35	49	111	492	5	30	42	55	106	486
Daze	0	7	11	16	32	501	0	9	15	21	48	503	0	17	22	28	45	497

Fourth Grade

DORF	7	75	103	128	210	589	0	94	118	142	213	596	8	106	130	154	226	589
DORFE	0	2	3	5	30	589	0	1	2	4	18	596	0	1	2	3	68	589

DORFA	.42	.94	.97	.98	1.00	589	.45	.96	.98	.99	1.00	592	.42	.97	.99	.99	1.00	589
Retell	0	21	33	47	94	577	0	25	35	50	101	592	2	30	40	55	114	582
Daze	0	12	17	21	50	585	0	13	18	24	48	589	0	22	26	32	49	586

Fifth Grade

DORF	6	96	125	149	243	546	0	109	133	156	256	543	26	116	143	167	280	539
DORFE	0	1	2	4	24	546	0	1	2	3	16	543	0	1	2	3	18	539
DORFA	.46	.96	.98	.99	1.00	546	.60	.97	.99	.99	1.00	542	.72	.98	.99	.99	1.00	539
Retell	0	27	37	50	108	541	0	30	40	57	113	541	7	31	42	59	129	530
Daze	0	15	20	25	53	543	0	15	20	26	44	537	0	21	27	33	57	536

Sixth Grade

DORF	12	109	137	154	219	535	9	111	138	162	242	528	16	122	146	171	246	531
DORFE	0	1	2	4	23	535	0	1	2	4	16	528	0	1	2	3	11	531
DORFA	.57	.97	.98	.99	1.00	535	.45	.97	.99	.99	1.00	528	.69	.98	.99	.99	1.00	531
Retell	0	29	40	53	98	531	0	31	46	61	98	526	3	35	50	65	109	527
Daze	0	18	22	27	50	535	0	20	26	33	63	520	0	20	27	33	64	529

Note. FSF = First Sound Fluency; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; DORF = DIBELS Oral Reading Fluency; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score.

Table 17

Descriptive Statistics for the Change in Scores between Times of Year on DIBELS Next Measures by Grade

Measure	Beginning to middle of year			Middle to end of year			Beginning to end of year		
	<i>Mean</i>	<i>SD</i>	N	<i>Mean</i>	<i>SD</i>	N	<i>Mean</i>	<i>SD</i>	N
Kindergarten									
FSF	20.27	12.69	467	-	-	-	-	-	-
LNF	19.76	13.65	467	8.87	11.88	466	26.63	15.76	458
PSF	-	-	-	13.05	18.21	466	-	-	-
CLS	-	-	-	13.19	15.51	464	-	-	-
WWR	-	-	-	4.14	6.45	462	-	-	-
First Grade									
CLS	20.94	18.51	451	16.88	19.16	447	37.89	22.45	441
WWR	8.62	8.28	450	5.97	8.21	446	14.60	10.32	441
DORF	-	-	-	21.26	14.76	445	-	-	-
DORFE	-	-	-	-1.38	4.13	444	-	-	-

DORFA	-	-	-	.09	.10	441	-	-	-
Retell	-	-	-	13.55	15.25	205	-	-	-
Second Grade									
DORF	18.59	14.79	548	15.32	13.87	555	33.79	17.86	540
DORFE	-1.52	3.70	547	-0.33	3.04	555	-1.85	3.57	539
DORFA	.04	.07	546	.02	.04	555	.06	.08	538
Retell	4.97	15.66	506	7.04	14.08	523	12.30	17.04	489
Third Grade									
DORF	14.25	13.62	492	12.52	13.78	490	26.75	15.46	483
DORFE	-1.14	2.57	492	0.14	2.49	490	-0.99	2.84	483
DORFA	.03	.05	492	.00	.03	490	.00	.03	490
Retell	7.65	15.05	469	4.84	15.47	475	12.33	16.03	459
Daze	4.05	5.72	490	6.54	6.34	488	10.62	5.92	480
Fourth Grade									
DORF	14.47	12.73	581	14.12	13.84	584	28.56	15.83	576
DORFE	-0.99	2.61	581	-0.41	3.56	584	-1.41	3.90	576
DORFA	.02	.04	581	.01	.03	584	.01	.03	584
Retell	2.80	15.84	568	5.90	15.75	577	8.62	18.13	561

Daze	1.82	5.29	572	7.96	5.53	575	9.78	5.90	571
Fifth Grade									
DORF	10.61	13.79	537	7.35	13.60	526	17.86	16.55	522
DORFE	-0.54	2.16	537	-0.25	1.80	526	-0.78	2.37	522
DORFA	.01	.03	537	.00	.02	526	.00	.02	526
Retell	4.65	15.73	531	1.04	15.76	518	5.50	16.82	511
Daze	-0.47	5.62	530	7.26	5.93	518	6.72	7.06	518
Sixth Grade									
DORF	3.70	13.84	518	8.64	12.78	521	12.21	13.51	518
DORFE	-0.10	2.25	518	-0.49	1.84	521	-0.59	2.29	518
DORFA	.00	.02	518	.01	.02	521	.01	.02	521
Retell	5.81	14.72	514	3.84	14.82	517	9.70	15.56	512
Daze	4.00	6.31	511	0.82	6.35	510	4.68	6.34	515

Note. FSF = First Sound Fluency; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WRC = Words Recoded Correctly; WUF = Word Use Fluency; DORF = DIBELS Oral Reading Fluency; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score.

Table 18

BGS DIBELS Composite Score Descriptive Statistics

[illegible]

Beginning	377.72	112.45	12	317	392	454	700	541
Middle	402.92	105.24	32	340	402	477	687	535
End	442.08	114.00	26	385	446	515	738	532
Sixth Grade								
Beginning	403.36	103.46	21	347	409	470	691	534
Middle	437.78	120.89	9	368	442	517	770	519
End	458.74	114.94	16	387	466	539	777	526

Note. N = 3,816. Students who did not have scores in all contributing measures were not given a composite score. DIBELS Composite Score is comprised of various measures across grades. Kindergarten: LNF and PSF at beginning of year; FSF, LNF, PSF, and NWF-CLS at middle of year; LNF, PSF, and NWF-CLS at end of year. First grade: LNF, PSF, and NWF-CLS at beginning of year; NWF-CLS, NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at middle of year; NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at end of year. Second grade: NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at beginning of year; Retell, DORF Words Correct, and a scaled score based on DORF Accuracy at middle of year and end of year. Third grade: Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year. Fourth, Fifth, and Sixth grades: Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year.

Reliability

Three forms of reliability were examined for the DIBELS measures at different benchmark periods with different districts. Inter-Rater reliability data were collected in the fall, and Alternate-Form and Test-Retest reliability data were collected in the winter.

Inter-rater reliability

All DIBELS Next measures were tested for reliability. Inter-rater reliability indicates the degree to which we get the same result when two independent examiners simultaneously score the same test during administration. Obtaining this information allows us to estimate the "extent to which we can generalize to different scorers" (Salvia & Ysseldyke, 2004, p. 126).

Descriptive statistics. Descriptive statistics for DIBELS Next scores by rater are reported in Tables 19 and 20. Mean scores across all grades differ by 1 point or less across raters (see Table 20). Differences in standard deviations are small and range from 0.05 to 1.3. Excluding Retell, the minimum, maximum, and median differ by no more than a two points. Quartile differences greater than 2 for Retell were found in fourth-, fifth-, and sixth-grade levels. The largest difference for the maximum score was in fifth grade (difference = 11). The largest difference for the minimum score was in sixth grade (difference = 7). We found the largest difference in the third quartile for fourth grade (difference = 4).

Correlations. Correlations for nearly all measures are above .90 and all correlations are significant at the $\alpha = .001$ level. (see Table 20). All but three correlations are at .85 or higher and most are .95 or higher. More specifically, measures with correlations at .99 or better are Daze, LNF, ORF, and both NWF scores. The correlation for PSF is .95. Retell inter-rater reliabilities are generally quite high, from .96 (fifth grade) to .99 (sixth grade) with the lowest Retell inter-rater reliability correlation in third grade at .92. The correlation for FSF is .94. The inter-rater reliability for ORF accuracy in third grade is .85. Examination of ORF scores and ORF Errors revealed a 10-point ORF Error discrepancy for one student; all other scores were within 1 point.

To more closely examine agreement between raters, we conducted paired t-tests under the equal variance and sample size assumption to test for differences in means between the two

rater's scores. The results suggest a difference in the two raters' scores for sixth grade Retell ($p = .04$). Further examination of Retell scores at this grade level revealed a 21-point scoring discrepancy for one student; all other pair-wise scores were within 2 points. For all other measures, we found no evidence of differences in the two rater's scores across all grades. These results suggest that for each measure, the mean difference between the two rater's scores was not significantly different from zero.

In general, inter-rater reliability for Retell Quality of Response is strong (Table 19). However, the inter-rater reliability for Retell Quality of Response is lower compared to the other scores. The median Quality of Response rating was 2 (provides 3 or fewer details) for all grades. Inter-rater reliability coefficients across all grades are about .69 or higher. Because this score is based on a 4-point scale and scoring can be subjective, a reliability coefficient and the percentage of exact-agreement between rater's above .50 is considered good and .75 is considered excellent. Across all grades, more than half of the raters agreed exactly on a student's Retell Quality of Response, and for 4 out of 5 grades, percent-agreement was greater than .75. Retell Quality of Response in fifth grade was the exception with a percent-agreement of .60.

Table 19

Inter-Rater Reliability for Retell Quality of Response

Quality of Response Rating by Rater											
Grade	N	First Rater				Second Rater				% - Agree	<i>r</i>
		1	2	3	4	1	2	3	4		
Second	19	5 (.26)	9 (.47)	5 (.26)	-	6 (.31)	9 (.47)	3 (.15)	1 (.05)	.84	.87
Third	22	4 (.18)	10 (.45)	8 (.36)	-	4 (.18)	9 (.41)	8 (.36)	1 (.05)	.77	.70
Fourth	23	4 (.17)	14 (.61)	5 (.22)	-	4 (.17)	16 (.70)	3 (.13)	1 (.05)	.91	.89
Fifth	28	9 (.32)	10 (.36)	7 (.25)	-	9 (.32)	15 (.54)	3 (.11)	1 (.04)	.60	.69
Sixth	20	4 (.20)	12 (.60)	2 (.10)	2 (.10)	3 (.15)	13 (.65)	2 (.10)	2 (.10)	.75	.81

Note. Percentage of students included in parenthesis. Quality of Response rating code: 1 = Provides 2 or fewer details; 2 = Provides 3 or fewer details; 3 = Provides 3 or more details in a meaningful sequence; 4 = Provides 3 or more details in a meaningful sequence with a main idea.

Table 20

Inter-Rater Reliability for DIBELS Measures

DIBELS Measures by Grade	N	First Rater				Second Rater				<i>r</i>
		<i>Mean</i>	<i>SD</i>	Min	Max	<i>Mean</i>	<i>SD</i>	Min	Max	
Kindergarten										
First Sound Fluency	25	12.36	11.98	0	35	11.56	12.17	0	36	.94
Letter Naming Fluency	25	20.52	14.31	0	46	20.12	14.50	0	46	.99
First Grade										
Letter Naming Fluency	25	48.52	19.79	0	99	48.68	19.90	0	99	.99
Phoneme Segmentation Fluency	25	38.76	17.16	0	63	37.20	16.29	0	61	.95
NWF: Correct Letter Sounds	25	41.32	32.18	9	139	40.80	32.41	9	140	.99
NWF: Whole Words Read	25	8.00	12.19	0	46	7.60	12.14	0	46	.99
Second Grade										
NWF: Correct Letter Sounds	25	64.08	32.63	18	140	64.00	33.39	20	142	.90
NWF: Whole Words Read	25	16.72	14.69	0	47	16.56	14.36	0	49	.99
Oral Reading Fluency Words Correct	25	58.72	28.67	15	106	58.32	29.37	14	108	.99
Oral Reading Fluency Accuracy	25	.90	.10	.68	1	.90	.10	.68	1	.99
Retell	20	26.60	12.65	0	56	26.75	13.35	0	58	.98

Retell Quality of Response	19	2.00	0.75	1	3	1.95	0.85	1	4	.87

Third Grade

Oral Reading Fluency Words Correct	25	95.24	37.97	14	193	94.68	37.79	15	192	.99
Oral Reading Fluency Accuracy	25	.95	.06	.70	.99	.95	.06	.75	1	.85
Retell	24	36.96	14.95	8	60	37.29	15.80	9	61	.92
Retell Quality of Response	22	2.18	0.73	1	3	2.27	0.83	1	4	.70
DIBELS Maze (Daze) raw scores	25	11.56	6.04	0	21	11.52	5.99	0	21	.99
Daze Errors	25	2.56	4.98	0	22	2.56	4.98	0	22	1.00
Daze Adjusted Score	25	10.60	6.64	0	21	10.56	6.60	0	21	.99

Fourth Grade

Oral Reading Fluency Words Correct	24	98.71	32.44	21	160	98.38	31.92	21	159	.99
Oral Reading Fluency Accuracy	24	.96	.04	.84	1	.96	.04	.84	1	.93
Retell	24	39.17	18.13	9	79	39.75	19.25	7	79	.98
Retell Quality of Response	23	2.04	0.64	1	3	1.96	0.56	1	3	.89
DIBELS Maze (Daze) raw scores	25	16.28	5.90	8	31	16.40	6.14	8	32	.99
Daze Errors	25	1.20	1.29	0	5	1.28	1.37	0	5	.96
Daze Adjusted Score	25	15.92	6.20	7	31	15.96	6.33	7	32	.98

Fifth Grade

Oral Reading Fluency Words Correct	28	110.04	40.24	35	184	110.25	40.51	38	186	.99
Oral Reading Fluency Accuracy	28	.96	.04	.85	1	.95	.04	.85	1	.95
Retell	28	35.79	16.96	0	59	35.07	18.26	0	70	.96
Retell Quality of Response	28	2.07	0.94	1	4	1.86	0.76	1	4	.69
DIBELS Maze (Daze) raw scores	26	22.31	9.26	5	40	22.23	9.29	5	40	.99
Daze Errors	25	2.32	2.43	0	8	2.40	2.40	0	8	.99
Daze Adjusted Score	26	20.81	9.87	1	40	21.23	9.95	1	40	.99
Sixth Grade										
Oral Reading Fluency Words Correct	20	140.80	32.30	86	219	141.25	32.34	85	219	.99
Oral Reading Fluency Accuracy	20	.98	.02	.92	1	.98	.02	.92	1	.91
Retell	20	41.10 ^a	19.60	13	90	42.50 ^a	19.31	13	90	.99
Retell Quality of Response	20	2.10	0.85	1	4	2.15	0.81	1	4	.81
DIBELS Maze (Daze) raw scores	20	22.65	8.54	4	44	22.50	8.69	4	44	.99
Daze Errors	20	0.50	0.69	0	2	0.50	0.69	0	2	1.00
Daze Adjusted Score	20	22.55	8.61	4	44	22.40	8.75	4	44	.99

Note. Correlations were significant at $p < .001$.

^aMeans are significantly different.

Alternate form reliability

Alternate-form reliability (sometimes called equivalent form reliability) is a means of generalizing to different item samples. Students are tested with two different, but "equivalent," forms of the test and scores from these two tests are correlated. Alternate form reliability data are reported in Table 21.

Descriptive Statistics. Mean scores across all grades differ by 8 points or less. Differences in standard deviations are less than 5 points for all scores with the exception of DORF Words Correct in sixth grade which differed by approximately 9 points.

Correlations. Correlations range from .44 to .99, excluding Daze errors, which ranges from .17 to .51. Measures that correlate higher than .90 are NWF WWR in first grade, DORF Words Correct for all grades, and Daze in third and sixth grade.

Test-retest reliability

Test-retest reliability is an index of score stability and indicates the degree to which we get the same result when a person is tested twice. The second test is given a short time after the first (Salvia & Ysseldyke say "preferably two weeks"). Scores from the two test administrations are then correlated. Test-retest reliability data are reported in Table 22.

Descriptive Statistics. Mean scores across all grades varied from form 1 to form 2 by a range of 2 to 19 points. Measures with a difference between scores of 5 or less were NWF WWR in kindergarten and DORF retell in second and third grade. We found a difference of 13 points between the test retest scores for PSF in kindergarten. DORF Words Correct scores had the largest difference (between 10 and 19 points) in second through sixth grades.

Correlations. Correlations range from .57 (DORF Accuracy) to .97 (DORF Words Correct), excluding DORF Retell which ranges from .27 to .74, and FSF in kindergarten which has a

correlation of .09. Measures that correlate higher than .90 are DORF Words Correct for all grades, and DORF accuracy in fourth through sixth grade.

Table 21

Alternate-Form Reliability for DIBELS Measures

DIBELS Measures by Grade	N	First Form		Second Form (AF)		<i>r</i>
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
Kindergarten						
First Sound Fluency	29	32.34	10.67	32.79	6.65	.52**
Letter Naming Fluency	29	39.76	15.90	45.48	15.64	.86***
Phoneme Segmentation Fluency	29	25.45	14.46	28.97	11.43	.44***
NWF: Correct Letter Sounds	27	17.37	10.78	21.89	14.82	.71***
NWF: Whole Words Read	27	0.74	1.81	2.04	3.78	.92***
First Grade						
NWF: Correct Letter Sounds	28	53.25	23.91	54.18	25.96	.85***
NWF: Whole Words Read	28	9.50	12.00	10.29	12.52	.90***
Oral Reading Fluency	28	37.39	40.06	46.00	41.57	.98***
Oral Reading Fluency Accuracy	28	.77	.15	.84	.11	.88***
Oral Reading Fluency Retell	11	23.55	12.48	22.09	16.23	.82**
Second Grade						
Oral Reading Fluency Words Correct	24	75.08	42.06	82.46	38.01	.97***

Oral Reading Fluency Accuracy	24	.91	.09	.93	.08	.83***
Oral Reading Fluency Retell	20	26.6	13.32	29.73	17.22	.68**

Third Grade

Oral Reading Fluency Words Correct	30	91.87	39.93	95.80	35.21	.96***
Oral Reading Fluency Accuracy	30	.96	.04	.95	.05	.80***
Oral Reading Fluency Retell	27	32.11	20.00	27.80	16.33	.81***
DIBELS Maze (Daze) raw score	24	11.92	7.94	15.25	7.30	.92***
Daze Errors	24	1.92	2.00	3.67	7.12	.18
Daze Adjusted Score	24	11.13	7.83	13.75	7.93	.86***

Fourth Grade

Oral Reading Fluency Words Correct	30	104.47	39.48	110.43	37.86	.96***
Oral Reading Fluency Accuracy	30	.96	.05	.97	.04	.85***
Oral Reading Fluency Retell	30	34.17	18.16	38.50	18.74	.80***
DIBELS Maze (Daze) raw score	29	16.93	5.92	21.48	7.41	.68***
Daze Errors	29	1.83	1.49	1.48	1.50	.39*
Daze Adjusted Score	29	16.34	5.92	20.93	7.28	.67***

Fifth Grade

Oral Reading Fluency Words Correct	25	113.56	27.96	120.48	27.98	.95***

Oral Reading Fluency Accuracy	25	.97	.02	.98	.02	.76***
Oral Reading Fluency Retell	25	37.24	15.86	36.04	18.55	.65***
DIBELS Maze (Daze) raw score	20	13.85	5.58	23.85	8.04	.45*
Daze Errors	20	2.00	1.78	1.70	1.38	.51*
Daze Adjusted Score	20	13.15	5.96	23.35	8.41	.49*

Sixth Grade

Oral Reading Fluency Words Correct	14	118.43	32.29	124.36	41.31	.96***
Oral Reading Fluency Accuracy	14	.96	.03	.96	.03	.74**
Oral Reading Fluency Retell	14	35.00	16.40	41.93	11.65	.64*
DIBELS Maze (Daze) raw score	6	29.17	12.32	32.00	13.11	.99***
Daze Errors	6	3.00	2.61	1.83	2.23	.17
Daze Adjusted Score	6	28.00	12.70	31.17	14.08	.98***

Note. Significance levels of correlations are denoted with the following: * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 22

Test-Retest Reliability for DIBELS Measures

DIBELS Measures by Grade	N	First Form		Second Form (TR)		<i>r</i>
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
Kindergarten						
First Sound Fluency	17	43.65	11.56	49.65	8.62	.09
Letter Naming Fluency	17	42.71	12.19	49.12	14.73	.83***
Phoneme Segmentation Fluency	17	41.24	20.74	54.24	11.94	.82***
NWF: Correct Letter Sounds	17	29.65	17.21	36.18	19.47	.72**
NWF: Whole Words Read	17	2.94	4.44	6.35	7.49	.67**
First Grade						
NWF: Correct Letter Sounds	27	58.63	22.27	69.00	22.83	.76***
NWF: Whole Words Read	27	12.63	10.58	17.11	11.54	.70***
Oral Reading Fluency Words Correct	28	35.86	26.22	44.29	28.66	.95***
Oral Reading Fluency Accuracy	28	0.77	0.22	0.83	0.14	.84***
Oral Reading Fluency Retell	11	24.82	15.32	31.73	16.74	.74**

Second Grade

Oral Reading Fluency Words Correct	21	102.38	27.74	113.76	28.37	.91***
Oral Reading Fluency Accuracy	21	0.97	0.03	0.99	0.01	.57**
Oral Reading Fluency Retell	21	48.33	15.21	49.86	17.81	.27

Third Grade

Oral Reading Fluency Words Correct	27	104.93	35.03	123.37	38.51	.93***
Oral Reading Fluency Accuracy	27	0.97	0.02	0.99	0.02	.68***
Oral Reading Fluency Retell	27	57.07	20.22	58.89	19.78	.69***

Fourth Grade

Oral Reading Fluency Words Correct	21	121.14	38.49	140.14	37.09	.97***
Oral Reading Fluency Accuracy	21	0.97	0.03	0.99	0.02	.91***
Oral Reading Fluency Retell	21	57.57	22.11	52.90	15.18	.36

Fifth Grade

Oral Reading Fluency Words Correct	23	124.43	42.71	134.13	43.56	.97***
Oral Reading Fluency Accuracy	23	0.96	0.05	0.97	0.08	.94***
Oral Reading Fluency Retell	22	52.32	19.15	60.27	15.75	.58**

Sixth Grade

Oral Reading Fluency Words Correct	15	144.47	42.8	157.67	42.23	.95***
Oral Reading Fluency Accuracy	15	0.98	0.02	0.98	0.03	.94***
Oral Reading Fluency Retell	15	61.47	18.06	67.87	24.25	.69**

Note. Descriptive statistics are unavailable for DIBELS Maze. Significance levels of correlations are denoted with the following: * $p < .05$; ** $p < .01$; *** $p < .001$. Reliability of Daze is not available.

DIBELS Next Correlations

Correlations between each measure by grade and time of year are presented in Tables 23 through 29. For kindergarten, the correlations for the same measure over time range from .41 for PSF at middle of year with PSF at end of year to .68 for NWF-CLS at middle of year with NWF-CLS at end of year (see Table 23). The highest correlation in kindergarten was for NWF-CLS at the end of the year with NWF-WWR at the end of the year. The correlations between the two phonemic awareness measures, FSF and PSF, range from .29 to .56. The lowest correlation is between these two measures given the farthest apart and the highest correlation is for concurrent assessments at middle of year.

First grade correlations for the same measures over time range from .53 for Retell at the middle of the year with Retell at the end of the year to .91 for DORF Words Correct middle of the year with DORF Words Correct at the end of the year (see Table 24). Strong correlations are observed between NWF-CLS and DORF Words Correct ranging from .77 to .80 and for NWF-WWR with DORF Words Correct ranging from .74 to .78. Moderate-strong correlations were observed for DORF Accuracy with NWF-CLS and DORF Accuracy with NWF-WWR, ranging from .57 to .64 and from .53 to .63 respectively.

The correlations for measures given in second grade are shown in Table 25. Correlations between DORF given at different times range from .89 to .93. For DORF accuracy the correlations range from .50 to .69 and for Retell they range from .50 to .69. When examining measures that assess the alphabetic principle and decoding skills, correlations range from .70 to .79 for DORF with NWF-CLS and from .66 to .74 for DORF with NWF-WWR. For DORF Accuracy, correlations with NWF-CLS and NWF-WWR were slightly lower; .44 to .56 and .41 to .53 respectively. Across all DIBELS Next measures at this grade level, lower correlations

were observed for assessments given farther apart in time and the highest correlations were found between middle of year and end of year assessments.

The correlations for third grade DORF given at the three time points are very high and range from .91 to .93 (see Table 26). Correlations for DORF Accuracy are also strong and range from .85 to .88. Daze correlations range from .72 to .76, while correlations for Retell were lower and ranged from .61 to .68. Correlations between Retell and Daze, both indicators of comprehension, range from .48 to .53.

Correlations for DIBELS Next measures given in fourth grade are reported in Table 27. Patterns of the correlations in fourth grade are similar to those observed in third grade. Correlations for DORF range from .91 to .94. DORF Accuracy correlations range from .78 to .89 and for Retell they range from .59 to .70. Correlations for Daze range from .75 to .78. Correlations for measures given at beginning of the year with the same measure given at the end of the year were lower than when comparing beginning of the year to middle- or end-of-year assessments. Correlations between Retell and Daze, both indicators of comprehension, range from .40 to .52.

DIBELS Next measure correlations for fifth-grade are reported in Table 28. As with third and fourth grade, correlations for DORF across time points are strong, ranging from .91 to .94. DORF Accuracy correlations range from .84 to .89 and Retell correlations range from .36 to .67. Daze correlations range from .72 to .80. For each of these measures, correlations were highest when examining scores from middle to end of year. When examining the data for Retell and Daze at each point in time, correlations range from .31 to .53. Daze at middle of year and Retell and the middle of year were the highest correlation observed.

Correlations for measures given in sixth grade are shown in Table 29. Once again, the highest correlations were for DORF and range from .93 to .94. Correlations for other measures were also quite strong with DORF Accuracy correlations ranging from .83 to .89, Retell from .70 to .75, and Daze ranging from .77 to .79. Correlations between Daze and Retell ranged from .40 to .52 (median of .42) with the highest correlations occurring between middle of year Retell and Daze at all three time points.

Overall, we noticed a steady increase in the correlation coefficients as grade level increases for all *DIBELS Next* measures with the exception of Retell. Correlations for Retell were the lowest in the upper grades. In contrast, correlations for DORF words correct were highest in the upper grades. All negative correlations between measures in first through sixth grade are the result of correlations with a measure of the number of errors occurring in DORF or DIBELS Maze. All correlations are significant at the $\alpha = .001$ level

Table 23

Correlations for DIBELS Kindergarten Measures

Measure - Time of Year	2	3	4	5	6
1. FSF - Beg.	.57 (478)	.47 (467)	.40 (467)	.47 (467)	.47 (467)
2. LNF - Beg.	-	.37 (467)	.63 (467)	.34 (467)	.56 (467)
3. FSF - Mid.		-	.50 (476)	.56 (476)	.46 (476)
4. LNF - Mid.			-	.44 (476)	.62 (476)
5. PSF - Mid.				-	.45 (476)
6. CLS - Mid.					-
	7	8	9	10	11
1. FSF - Beg.	.36 (464)	.31 (458)	.29 (458)	.41 (457)	.35 (457)
2. LNF - Beg.	.39 (464)	.49 (458)	.18 (458)	.49 (457)	.39 (457)
3. FSF - Mid.	.26 (474)	.45 (466)	.42 (466)	.35 (464)	.25 (464)
4. LNF - Mid.	.41 (474)	.72 (466)	.27 (466)	.49 (464)	.43 (464)
5. PSF - Mid.	.26 (474)	.41 (466)	.41 (466)	.37 (464)	.31 (464)
6. CLS - Mid.	.64 (474)	.55 (466)	.33 (466)	.68 (464)	.61 (464)
7. WWR - Mid.	-	.35 (464)	.22 (464)	.62 (462)	.67 (462)
8. LNF - End		-	.47 (479)	.62 (477)	.53 (477)
9. PSF - End			-	.43 (477)	.35 (477)
10. CLS - End				-	.84 (477)
11. WWR - End					-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. FSF = First Sound Fluency; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All

correlations are significant at $p < .001$.

Table 24

Correlations for DIBELS First Grade Measures

Measure - Time of Year	2	3	4	5	6
1. LNF - Beg.	.37 (461)	.61 (461)	.51 (461)	.57 (451)	.55 (450)
2. PSF - Beg.	-	.30 (461)	.18 (461)	.24 (451)	.19 (450)
3. CLS - Beg.		-	.90 (461)	.83 (451)	.79 (450)
4. WWR - Beg.			-	.79 (451)	.79 (450)
5. CLS - Mid.				-	.93 (457)
6. WWR - Mid.					-
	7	8	9	10	11
1. LNF - Beg.	.64 (451)	-.32 (451)	.58 (446)	.55 (264)	.58 (441)
2. PSF - Beg.	.24 (451)	-.14 (451)	.29 (446)	.16 (264)	.24 (441)
3. CLS - Beg.	.80 (451)	-.46 (451)	.57 (446)	.63 (264)	.75 (441)
4. WWR - Beg.	.77 (451)	-.49 (451)	.53 (446)	.56 (264)	.71 (441)
5. CLS - Mid.	.80 (458)	-.49 (457)	.64 (451)	.65 (271)	.83 (447)
6. WWR - Mid.	.78 (457)	-.50 (456)	.63 (450)	.58 (270)	.81 (446)
7. DORF - Mid.	-	-.55 (457)	.74 (451)	.76 (271)	.73 (447)
8. DORFE - Mid.		-	-.77 (451)	-.39 (270)	-.44 (446)
9. DORFA - Mid.			-	.56 (265)	.62 (442)
10. Retell - Mid.				-	.59 (264)
11. CLS - End					-
	12	13	14	15	16

1. LNF - Beg.	.54 (441)	.66 (439)	-.35 (439)	.52 (439)	.28 (358)
2. PSF - Beg.	.20 (441)	.21 (439)	-.22 (439)	.30 (439)	.04† (358)
3. CLS - Beg.	.71 (441)	.70 (439)	-.40 (439)	.45 (439)	.35 (358)
4. WWR - Beg.	.70 (441)	.66 (439)	-.38 (439)	.39 (439)	.27 (358)
5. CLS - Mid.	.81 (447)	.74 (445)	-.42 (445)	.51 (445)	.34 (361)
6. WWR - Mid.	.83 (446)	.72 (444)	-.43 (444)	.50 (444)	.30 (360)
7. DORF - Mid.	.71 (447)	.91 (445)	-.42 (445)	.55 (445)	.39 (361)
8. DORFE - Mid.	-.46 (446)	-.52 (444)	.40 (444)	-.45 (444)	-.26 (360)
9. DORFA - Mid.	.61 (442)	.78 (441)	-.54 (441)	.79 (441)	.31 (360)
10. Retell - Mid.	.57 (264)	.73 (262)	-.32 (262)	.44 (262)	.53 (205)
11. CLS - End	.94 (461)	.77 (459)	-.45 (459)	.57 (459)	.28 (373)
12. WWR - End	-	.74 (459)	-.48 (459)	.58 (459)	.25 (373)
13. DORF - End		-	-.49 (459)	.70 (459)	.44 (373)
14. DORFE - End			-	-.78 (459)	-.17 (373)
15. DORFA - End				-	.29 (373)
16. Retell - End					-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. Unless otherwise noted, correlations are significant at $p < .001$.

† Not significant.

Table 25

Correlations for DIBELS Second Grade Measures

Measure - Time of Year	2	3	4	5	6	7	
1. CLS - Beg.	.92 (560)	.79 (560)	-.39 (559)	.56 (558)	.50 (520)	.73 (548)	
2. WWR - Beg.	-	.74 (560)	-.41 (559)	.53 (558)	.44 (520)	.69 (548)	
3. DORF - Beg.		-	-.46 (559)	.69 (558)	.61 (520)	.92 (548)	
4. DORFE - Beg.			-	-.76 (558)	-.29 (520)	-.46 (547)	
5. DORFA - Beg.				-	.47 (519)	.70 (546)	
6. Retell - Beg.					-	.62 (509)	
7. DORF - Mid.						-	
	8	9	10	11	12	13	14
1. CLS - Beg.	-.50 (548)	.52 (548)	.47 (531)	.70 (540)	-.41 (540)	.44 (540)	.42 (519)
2. WWR - Beg.	-.48 (548)	.48 (548)	.43 (531)	.66 (540)	-.42 (540)	.41 (540)	.39 (519)
3. DORF - Beg.	-.55 (548)	.61 (548)	.57 (531)	.89 (540)	-.49 (540)	.53 (540)	.53 (519)
4. DORFE - Beg.	.46 (547)	-.45 (547)	-.20 (531)	-.40 (539)	.45 (539)	-.38 (539)	-.17 (519)
5. DORFA - Beg.	-.60 (546)	.82 (548)	.44 (530)	.67 (538)	-.59 (538)	.73 (538)	.36 (519)

6. Retell - Beg.	-.34 (509)	.43 (509)	.56 (506)	.63 (502)	-.27 (502)	.38 (502)	.50 (489)
7. DORF - Mid.	-.55 (457)	.70 (564)	.76 (271)	.93 (555)	-.53 (555)	.59 (555)	.57 (533)
8. DORFE - Mid.	-	-.82 (564)	-.55 (457)	-.58 (555)	.56 (555)	-.58 (555)	-.34 (533)
9. DORFA - Mid.		-	.48 (547)	.69 (555)	-.65 (555)	.86 (555)	.41 (533)
10. Retell - Mid.			-	.62 (538)	-.31 (538)	.39 (538)	.69 (523)
11. DORF - End				-	-.55 (566)	.65 (566)	.61 (541)
12. DORFE - End					-	-.80 (566)	-.26 (541)
13. DORFA - End						-	.33 (541)
14. Retell - End							-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; CLS = Correct Letter Sounds; WWR = Whole Words Read; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All correlations significant at $p < .001$.

Table 26

Correlations for DIBELS Third Grade Measures

Measure - Time of Year	2	3	4	5	6	7	8

1. DORF - Beg.	-.50 (502)	.66 (502)	.61 (479)	.77 (501)	.93 (492)	-.54 (492)	.62 (492)
2. DORFE - Beg.	-	-.77 (502)	-.24 (479)	-.44 (501)	-.51 (492)	.60 (492)	-.59 (492)
3. DORFA - Beg.		-	.40 (479)	.51 (501)	.66 (492)	-.62 (492)	.88 (492)
4. Retell - Beg.			-	.51 (478)	.55 (470)	-.24 (470)	.36 (470)
5. Daze - Beg.					.74 (491)	-.44 (491)	.47 (491)
6. DORF - Mid.						-.53 (504)	.66 (502)
7. DORFE - Mid.						-	-.78 (502)
8. DORFA - Mid.							-
	9	10	11	12	13	14	15
1. DORF - Beg.	.56 (480)	.70 (491)	.91 (483)	-.51 (483)	.58 (483)	.50 (473)	.78 (481)
2. DORFE - Beg.	-.25 (480)	-.42 (491)	-.48 (483)	.56 (483)	-.54 (483)	-.23 (473)	-.43 (481)
3. DORFA - Beg.	.37 (480)	.51 (491)	.65 (483)	-.60 (483)	.86 (483)	.33 (473)	.58 (481)
4. Retell - Beg.	.67 (469)	.51 (469)	.57 (461)	-.18 (461)	.29 (461)	.61 (459)	.53 (460)
5. Daze - Beg.	.52 (479)	.76 (490)	.72 (482)	-.43 (482)	.46 (482)	.48 (472)	.72 (480)
6. DORF - Mid.	.57 (492)	.69 (503)	.93 (490)	-.51 (490)	.60 (490)	.47 (479)	.78 (489)
7. DORFE - Mid.	-.25 (492)	-.41 (503)	-.55 (490)	.57 (490)	-.57 (490)	-.20 (479)	-.45 (489)
8. DORFA - Mid.	.38 (490)	.48 (501)	.65 (490)	-.58 (490)	.85 (490)	.33 (479)	.56 (489)
9. Retell - Mid.	-	.52 (491)	.54 (478)	-.25 (478)	.34 (478)	.68 (475)	.51 (478)

10. Daze - Mid.	.67 (489)	-.39 (489)	.45 (489)	.48 (478)	.73 (488)
11. DORF - End	-	-.53 (498)	.64 (498)	.52 (486)	.78 (496)
12. DORFE - End		-	-.80 (498)	-.24 (486)	-.46 (496)
13. DORFA - End			-	.33 (486)	.55 (496)
14. Retell - End				-	.48 (486)
15. Daze - End					-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All correlations are significant at $p < .001$.

Table 27

Correlations for DIBELS Fourth Grade Measures

Measure - Time of Year	2	3	4	5	6	7	8
1. DORF - Beg.	-.49 (589)	.67 (589)	.57 (577)	.78 (584)	.94 (581)	-.54 (581)	.63 (581)
2. DORFE - Beg.	-	-.77 (589)	-.20 (577)	-.43 (584)	-.51 (581)	.63 (581)	-.57 (581)
3. DORFA - Beg.		-	.39 (577)	.56 (584)	.69 (581)	-.70 (581)	.89 (581)

4. Retell - Beg.			-	.45 (572)	.54 (569)	-.22 (569)	.35 (569)
5. Daze - Beg.				-	.75 (579)	-.45 (579)	.51 (579)
6. DORF - Mid.					-	-.53 (596)	.68 (592)
7. DORFE - Mid.						-	-.82 (592)
8. DORFA - Mid.							-
	9	10	11	12	13	14	15
1. DORF - Beg.	.58 (577)	.78 (574)	.91 (576)	-.27 (576)	.51 (576)	.52 (570)	.77 (574)
2. DORFE - Beg.	-.30 (577)	-.37 (574)	-.49 (576)	.34 (576)	-.49 (576)	-.24 (570)	-.41 (574)
3. DORFA - Beg.	.42 (577)	.51 (574)	.67 (576)	-.37 (576)	.78 (576)	.35 (570)	.61 (574)
4. Retell - Beg.	.65 (568)	.47 (562)	.51 (564)	-.07† (564)	.26 (564)	.59 (561)	.48 (562)
5. Daze - Beg.	.49 (575)	.76 (572)	.74 (574)	-.27 (574)	.46 (574)	.45 (568)	.75 (571)
6. DORF - Mid.	.60 (592)	.79 (589)	.93 (584)	-.27 (584)	.52 (584)	.49 (577)	.79 (581)
7. DORFE - Mid.	-.24 (517)	-.37 (589)	-.53 (584)	.37 (584)	-.52 (584)	-.21 (577)	-.43 (581)
8. DORFA - Mid.	.39 (588)	.49 (585)	.65 (584)	-.38 (584)	.81 (584)	.32 (577)	.58 (581)
9. Retell - Mid.	-	.50 (585)	.56 (580)	-.09* (580)	.29 (580)	.70 (577)	.52 (577)
10. Daze - Mid.		-	.76 (577)	-.22 (577)	.41 (577)	.40 (570)	.78 (575)
11. DORF - End			-	-.30 (589)	.56 (589)	.52 (582)	.78 (586)
12. DORFE - End				-	-.67 (589)	-.14 (582)	-.26 (586)

13. DORFA - End	-	.28 (582)	.49 (586)
14. Retell - End		-	.45 (579)
15. Daze - End			-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All correlations are significant at $p < .001$. † Not significant.

Table 28

Correlations for DIBELS Fifth Grade Measures

Measure - Time of Year	2	3	4	5	6	7	8
1. DORF - Beg.	-.59 (546)	.66 (546)	.51 (541)	.74 (543)	.93 (537)	-.60 (537)	.62 (537)
2. DORFE - Beg.	-	-.82 (546)	-.34 (541)	-.49 (543)	-.58 (537)	.68 (537)	-.69 (537)
3. DORFA - Beg.		-	.38 (541)	.52 (543)	.64 (537)	-.71 (537)	.89 (537)
4. Retell - Beg.			-	.42 (538)	.46 (532)	-.32 (532)	.36 (532)
5. Daze - Beg.				-	.71 (535)	-.49 (535)	.49 (535)
6. DORF - Mid.					-	-.58 (543)	.64 (542)

7. DORFE - Mid. - -.87 (542)

8. DORFA - Mid. -

	9	10	11	12	13	14	15
1. DORF - Beg.	.62 (537)	.78 (531)	.91 (522)	-.54 (522)	.62 (522)	.44 (514)	.78 (520)
2. DORFE - Beg.	-.69 (537)	-.49 (531)	-.58 (522)	.62 (522)	-.67 (522)	-.27 (514)	-.48 (520)
3. DORFA - Beg.	.89 (537)	.54 (531)	.63 (522)	-.64 (522)	.84 (522)	.30 (514)	.53 (520)
4. Retell - Beg.	.36 (532)	.42 (526)	.46 (517)	-.32 (517)	.36 (517)	.60 (511)	.42 (515)
5. Daze - Beg.	.49 (535)	.78 (530)	.71 (520)	-.46 (520)	.51 (520)	.31 (513)	.72 (518)
6. DORF - Mid.	.61 (541)	.78 (537)	.94 (526)	-.53 (526)	.63 (526)	.44 (518)	.79 (524)
7. DORFE - Mid.	-.36 (541)	-.51 (537)	-.60 (526)	.67 (526)	-.72 (526)	-.28 (518)	-.51 (524)
8. DORFA - Mid.	.40 (540)	.53 (536)	.63 (526)	-.67 (526)	.86 (526)	.30 (518)	.52 (524)
9. Retell - Mid.	-	.52 (536)	.57 (524)	-.32 (524)	.38 (524)	.67 (518)	.49 (522)
10. Daze - Mid.		-	.76 (520)	-.47 (520)	.53 (520)	.40 (513)	.80 (518)
11. DORF - End			-	-.57 (539)	.66 (539)	.45 (530)	.77 (536)
12. DORFE - End				-	-.90 (539)	-.29 (530)	-.48 (536)
13. DORFA - End					-	.33 (530)	.53 (536)
14. Retell - End						-	.43 (528)
15. Daze - End							-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All correlations are significant at $p < .001$.

Table 29

Correlations for DIBELS Sixth Grade Measures

Measure - Time of Year	2	3	4	5	6	7	8
1. DORF - Beg.	-.48 (535)	.66 (535)	.49 (531)	.77 (535)	.93 (518)	-.50 (518)	.61 (518)
2. DORFE - Beg.	-	-.77 (535)	-.27 (531)	-.41 (535)	-.46 (518)	.61 (518)	-.54 (518)
3. DORFA - Beg.		-	.35 (531)	.54 (535)	.62 (518)	-.63 (518)	.89 (518)
4. Retell - Beg.			-	.41 (531)	.50 (514)	-.23 (514)	.32 (514)
7. Daze - Beg.				-	.77 (518)	-.47 (518)	.52 (518)
8. DORF - Mid.					-	-.53 (528)	.62 (528)
9. DORFE - Mid.						-	-.76 (528)
10. DORFA - Mid.							-
	9	10	11	12	13	14	15
1. DORF - Beg.	.57 (516)	.76 (511)	.93 (518)	-.53 (518)	.64 (518)	.48 (514)	.75 (515)

2. DORFE - Beg.	-.32 (516)	-.43 (511)	-.43 (518)	.56 (518)	-.52 (518)	-.24 (514)	-.41 (515)
3. DORFA - Beg.	.38 (516)	.54 (511)	.62 (518)	-.59 (518)	.83 (518)	.31 (514)	.53 (515)
4. Retell - Beg.	.73 (514)	.40 (507)	.49 (514)	-.23 (514)	.31 (514)	.70 (512)	.42 (511)
5. Daze - Beg.	.51 (516)	.77 (511)	.78 (518)	-.47 (518)	.54 (518)	.42 (514)	.76 (515)
6. DORF - Mid.	.62 (526)	.78 (519)	.94 (521)	-.51 (521)	.61 (521)	.51 (517)	.78 (517)
7. DORFE - Mid.	-.33 (522)	-.54 (519)	-.48 (521)	.66 (521)	-.62 (521)	-.21 (517)	-.44 (517)
8. DORFA - Mid.	.40 (526)	.55 (519)	.60 (521)	-.59 (521)	.87 (521)	.30 (517)	.51 (517)
9. Retell - Mid.	-	.51 (517)	.59 (519)	-.34 (519)	.38 (519)	.75 (517)	.52 (515)
10. Daze - Mid.		-	.76 (512)	-.48 (512)	.53 (512)	.41 (508)	.79 (510)
11. DORF - End			-	-.51 (531)	.65 (531)	.53 (527)	.78 (527)
12. DORFE - End				-	-.79 (531)	-.27 (527)	-.44 (527)
13. DORFA - End					-	.33 (527)	.53 (527)
14. Retell - End						-	.46 (523)
15. Daze - End							-

Note. Correlations are based on subjects with pair-wise complete data. The number with pair-wise complete data is reported in parentheses. DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; Daze = Daze Adjusted Score; Beg. = Beginning of year; Mid. = Middle of year; End = End of year. All correlations are significant at $p < .001$.

Concurrent and Predictive Validity of DIBELS Next Measures

Criterion-related validity is the extent to which a person's performance on a criterion measure can be estimated from that person's performance on the assessment procedure being validated (Salvia, Ysseldyke, & Bolt, 2007). Concurrent validity describes the relationship between a student's performance on the assessment being studied and the criterion measure when both are given at approximately the same time, while predictive validity is how well student performance on the assessment predicts future performance on the criterion measure. In this section we present correlational evidence of criterion-related validity for the individual DIBELS measures with an external criterion measure, the GRADE. In addition, we also present data for an internal criterion, the DIBELS Composite Score (DCS).

The GRADE was administered during the spring near the end-of-year benchmark assessment. Thus, correlations with beginning and middle-of-year DIBELS scores indicate the predictive validity of DIBELS, and correlations with end-of-year scores represent the concurrent validity of DIBELS. The correlation associated with GRADE Total Test score, a composite of all grade-level subtests, provides an overview of the relationship between DIBELS and the GRADE-measured reading skill-level. The GRADE Total Test score is comprised of scores across subtests of the GRADE that vary by grade level. In kindergarten, the GRADE Total Test score is comprised of measures that assess phonics, phonemic and phonological awareness. In first and second grade, GRADE Total Test includes word meaning, sentence or passage reading, and comprehension measures. In third grade, GRADE Total Test is comprised of measures assessing word reading, vocabulary, and comprehension. In fourth, fifth, and sixth grade, GRADE Total Test includes scores from measures of vocabulary and comprehension. We reference the correlation with GRADE Total Test Score as the primary indicator of criterion-

related validity because it contains the most information about the student's overall reading skill with respect to GRADE results. Thus, these results are presented and discussed first, followed by the results with respect to other correlations of interest (i.e., correlations of DIBELS Next measures with GRADE composites or subtests of the same construct). All correlations of DIBELS Next Measures and the GRADE are presented by grade level in Tables 30 through 36. We present correlations for DIBELS Next measures with GRADE raw and standard scores, as well as stanines. However, correlations based upon raw scores are not appreciably different from, and in many cases identical to, those for standard scores or stanines.

Correlations of DIBELS Next Measures and GRADE Total Test. Most correlations for kindergarten DIBELS Next measures with GRADE Total Test Composite are in the moderate to moderate-strong range based on criteria delineated by Hopkins, 2002 (see Table 30). The exceptions were for WWR at mid-year and PSF at the end of the year, both of which were in the small range. First grade correlations varied more by time of year (see Table 31). At the beginning of the year, correlations for first-grade DIBELS Next measures with GRADE Total Test were moderate to moderate-strong with LNF having the highest correlation. At the middle of the year assessment, correlations ranged from moderate-strong to strong with DORF Accuracy being the highest correlation with GRADE Total Test ($r = .80$). At the end of first grade, the correlations were mostly moderate-strong to strong with DORF Words Correct and DORF Accuracy consistently showing the highest correlations.

Across second grade, correlations of DIBELS Next measures with GRADE Total Test ranged from moderate-strong to strong with the exception of Retell Quality of Response (RQR; see Table 32). For third grade DIBELS Next measures, most correlations with GRADE Total Test were in the moderate-strong range across all times of year (see Table 33). This pattern

changed slightly in fourth grade, where DORF Words Correct had consistently strong correlations with GRADE Total Test. Most of the other DIBELS Next measures had moderate-strong correlations (see Table 34). Both fifth- and sixth-grade results indicate mostly moderate-strong correlations across measures and times of year (see Tables 35 & 36). Overall, when examining individual measures, predictive and concurrent validity coefficients are moderate to strong for second- through sixth-grade DIBELS Next measures with the GRADE Total Test.

Correlations of DIBELS Next Measures and GRADE composites and subtests. In addition to the primary correlations of interest (i.e., DIBELS Next measures and their relation to each other, to the DCS, and to GRADE Total Test), we also examined correlations between DIBELS Next measures and composites/subtests of the GRADE measuring the same construct, or related construct of interest (e.g., vocabulary). The following GRADE subtests were examined: Phonological Awareness (PA) composite, Vocabulary (VOC) composite, Comprehension (COMP) composite, and Early Literacy Skills (ELS) composite, Phoneme-Grapheme Correspondence (PGC) subtest. Correlations are reported in Tables 30 - 36. Correlations based on raw scores are not appreciably different from, and in many cases identical to, those from standard or stanine scores, thus correlations based on raw test scores are discussed. Results are discussed by skill area/construct, beginning with phonemic awareness.

The two DIBELS Next measures of phonemic awareness are FSF and PSF. Correlations of FSF and PSF with the GRADE PA composite were examined in Kindergarten only because the subtests that comprise the PA composite are administered only in Kindergarten. These correlations range from small, but significant ($r = .16$, end-of-year PSF), to moderately-strong ($r = .50$, beginning-of-year FSF) (see Table 30). We also compared GRADE ELS and DIBELS LNF at Kindergarten only because Kindergarten is the only grade level at which the subtests that

comprise the ELS are administered. All correlations at each time of year between LNF and ELS were small.

The correlation between NWF-CLS and GRADE ELS composite in Kindergarten was small at middle- and end-of-year. However, the correlation between GRADE PGC subtest and NWF-CLS was moderate at both middle- and end-of-year. In first grade, NWF-CLS was correlated moderately with the GRADE Comprehension composite at the beginning of the year, but the correlations increased to moderate-strong at middle- and end-of year. At the beginning of second grade, the correlation between the GRADE COMP and NWF-CLS also was moderately-strong.

The correlations between NWF-WWR and GRADE ELS composite , as well as with the PGC subtest in kindergarten were small at both middle- and end-of-year. In first grade, NWF-WWR was correlated moderately with the GRADE Comprehension composite at the beginning of the year, but the correlations increased to moderate-strong at middle- and end-of year. At the beginning of second grade, the correlation between the GRADE COMP and NWF-WWR also was moderately-strong. In general, correlational relationships for NWF-CLS are lower for kindergarten and increase across first and second grades and time of year.

DORF Words Correct was examined in relation to the GRADE Comprehension and Vocabulary composites beginning at the middle of first grade and continuing up through the end of sixth grade (see Tables 31 - 36). The correlation of DORF at first grade middle-of-year was moderate with GRADE VOC and moderate-strong with GRADE COMP. At first grade end-of-year, these correlations increased to moderate-strong for VOC and strong for COMP. At the beginning of second grade, correlations of DORF with both the COMP and VOC composites were moderate-strong. At middle- and end-of-year DORF correlations with VOC were moderate-strong and strong with COMP. For third grade, the correlations of DORF with both COMP and

VOC were moderate-strong at all three time points. In fourth grade, DORF correlations increased to strong at all three time points for both COMP and VOC. However, at all time points in fifth and sixth grade, DORF correlations with COMP and VOC were moderate-strong. Overall, these DORF correlations suggest excellent concurrent and predictive validity with the GRADE Vocabulary and Comprehension composites.

The correlations for DORF Accuracy are reported in Tables 31-36. All DORF Accuracy correlations with COMP and VOC composites fell in the moderate-strong to strong range regardless of grade level. In middle-of-year first grade DORF Accuracy is strongly correlated with both COMP and VOC. At the end of the year, the correlations were strong with VOC and moderate-strong with COMP. At the beginning of second grade, the relation changed slightly with DORF Accuracy having a strong correlation with COMP and a moderate-strong correlation with VOC. However, at middle- and end-of-year the correlation with VOC was strong, while the correlation with COMP was strong at middle-of-year but moderate-strong at the end of the year. Similar small shifts in the strength of the relationship occurred in third grade. At the beginning- and end-of-year, the correlations of DORF Accuracy with VOC and COMP were moderate-strong. However, at middle-of-year, the correlations were moderate-strong with COMP and strong with VOC. At fourth grade, correlations remained consistent throughout the year and were all in the moderate-strong range with both VOC and COMP. In fifth grade the correlations at beginning-of-year were moderate-strong for both VOC and COMP, but at middle- and end-of-year, DORF Accuracy correlations with VOC and COMP were in the moderate range. Finally, for sixth grade at both the beginning- and end-of-year DORF Accuracy had moderate-strong correlations with both VOC and COMP, but at middle-of-year the correlations were lower and in

the moderate range. Overall, these correlations suggest excellent concurrent and predictive validity for DORF Accuracy with respect to the GRADE.

Correlations of Retell with GRADE VOC and COMP are reported in Tables 31 through 36. In first grade the correlations shift slightly from middle-of-year, when they are moderate with VOC and moderate-strong with COMP, to end-of-year, when they are small with VOC and moderate with COMP. In second grade, at both beginning- and middle-of-year the correlations of Retell with VOC were moderate and moderate-strong with COMP. Correlations were slightly lower at the end of second grade with Retell showing moderate correlations with both VOC and COMP. In third grade, the correlation between Retell and VOC was moderate at all three time points, but the correlation of Retell with COMP was moderate-strong at all three time points. Slightly higher correlations were observed in fourth and fifth grade. At both grade levels at all three times of year, correlations of Retell with both VOC and COMP were moderate-strong. A slightly lower correlation (moderate) was noted at the beginning of sixth grade between Retell and COMP. However, the correlations of Retell and COMP at middle- and end-of-year were moderate-strong as were the correlations of Retell and VOC at all three time points.

Correlations for Daze are available for third through sixth grade (the grade levels for which Daze is administered) (see Tables 33 - 36). For third, fourth and fifth grade, correlations of Daze with both VOC and COMP were moderate-strong for all three times of year. A slightly lower correlation (i.e., moderate) was found at middle-of-year in sixth grade for Daze with COMP, but a moderate-strong correlation between Daze and VOC was found. At beginning- and end-of-year the correlations were moderate-strong for both VOC and COMP in sixth grade. Overall, these correlations for Daze suggest very good concurrent and predictive validity with respect to the GRADE.

Table 30

Kindergarten GRADE Subtest and Composite Score Correlations with All DIBELS Measures

Subtests and Composites	FSF	LNF	PSF	CLS	WWR
Beginning of year					
Raw Scores					
1. Total Test	.52	.39	-	-	-
2. Phonological Awareness	.50	.37	-	-	-
3. Sound Matching	.45	.30	-	-	-
4. Rhyming	.41	.35	-	-	-
5. Early Literacy Skills	.32	.26**	-	-	-
6. Print Awareness	.26**	.29	-	-	-
7. Letter Recognition	.20**	.05†	-	-	-
8. Same & Different Words	.25	.30	-	-	-
9. Phoneme-Grapheme Correspondence	.32	.30	-	-	-
10. Listening Comprehension	.28	.21**	-	-	-
Standard Scores					
11. Total Test	.57	.45	-	-	-
Stanine Scores					
12. Total Test	.57	.46	-	-	-
13. Phonological Awareness	.50	.39	-	-	-
14. Early Literacy Skills	.37	.39	-	-	-
15. Phoneme-Grapheme Correspondence	.37	.37	-	-	-
16. Listening Comprehension	.36	.26	-	-	-
Middle of year					
Raw Scores					
1. Total Test	.40	.35	.34	.47	.19*

2. Phonological Awareness	.36	.30	.30	.38	.17*
3. Sound Matching	.36	.26	.31	.37	.19*
4. Rhyming	.26	.26	.20***	.30	.11†
5. Early Literacy Skills	.18*	.23***	.18*	.29	.18*
6. Print Awareness	.07†	.15*	.25	.25	.16*
7. Letter Recognition	.07†	.06†	.07†	.10†	.09†
8. Same & Different Words	.28	.31	.12†	.28	.12†
9. Phoneme-Grapheme Correspondence	.36	.34	.29	.39	.14†
10. Listening Comprehension	.26	.24	.27	.37	.13†

Standard Scores

11. Total Test	.41	.35	.38	.50	.20

Stanine Scores

12. Total Test	.40	.36	.39	.48	.19*
13. Phonological Awareness	.37	.27	.32	.40	.19
14. Early Literacy Skills	.20	.30	.24***	.36	.19*
15. Phoneme-Grapheme Correspondence	.39	.41	.33	.45	.22***
16. Listening Comprehension	.26	.30	.35	.38	.13†

End of year

Raw Scores

1. Total Test	-	.35	.24	.40	.35
2. Phonological Awareness	-	.28	.16*	.35	.30
3. Sound Matching	-	.33	.22***	.40	.40
4. Rhyming	-	.15*	.05†	.21***	.11†
5. Early Literacy Skills	-	.15*	.07†	.17*	.23***
6. Print Awareness	-	.12†	.08†	.20**	.16*
7. Letter Recognition	-	.06†	.06†	.03†	.17*

8. Same & Different Words	-	.16*	.00†	.17*	.14*
9. Phoneme-Grapheme Correspondence	-	.39	.29	.30	.27
10. Listening Comprehension	-	.28	.29	.34	.27
Standard Scores					
11. Total Test		.35	.25***	.49	.41
Stanine Scores					
12. Total Test	-	.35	.24***	.47	.40
13. Phonological Awareness	-	.28	.18*	.41	.37
14. Early Literacy Skills	-	.23***	.12†	.30	.23***
15. Phoneme-Grapheme Correspondence	-	.44	.26	.34	.32
16. Listening Comprehension	-	.29	.31	.34	.30

Note. The GRADE was administered during the end-of-year benchmark assessment. Pair-wise sample sizes for all times of year ≈ 170 . FSF = First Sound Fluency; LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read. Unless otherwise marked, all correlations are significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† Not significant.

Table 31

First Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

[illegible]

12. Total Test	.55	.31	.47	.43	-	-	-	-
13. Vocabulary Composite	.56	.25	.48	.45	-	-	-	-
14. Word Reading	.53	.24	.42	.39	-	-	-	-
15. Word Meaning	.43	.27	.40	.37	-	-	-	-
16. Comprehension Composite	.52	.28	.46	.42	-	-	-	-
17. Sentence Comprehension	.51	.26	.45	.42	-	-	-	-
18. Passage Comprehension	.44	.25	.37	.34	-	-	-	-
19. Listening Comprehension	.01*	.14*	.20**	.14	-	-	-	-

Middle of year

Raw Scores

1. Total Test	-	-	.51	.52	.64	-.59	.80	.55
2. Vocabulary Composite	-	-	.41	.41	.48	-.51	.70	.38
3. Word Reading	-	-	.36	.37	.40	-.33	.53	.27***
4. Word Meaning	-	-	.36	.36	.44	-.54	.70	.40
5. Comprehension Composite	-	-	.51	.51	.66	-.56	.77	.58
6. Sentence Comprehension	-	-	.51	.52	.60	-.54	.75	.52
7. Passage Comprehension	-	-	.45	.44	.61	-.50	.67	.54
8. Listening Comprehension	-	-	.22***	.21***	.17*	-.13†	.18*	.08†

Standard Scores								
9. Total Test	-	-	.58	.57	.72	-.60	.78	.61
10. Vocabulary Composite	-	-	.55	.55	.63	-.55	.72	.54
11. Comprehension Composite	-	-	.56	.55	.72	-.57	.75	.61
Stanine Scores								
12. Total Test	-	-	.56	.56	.69	-.60	.79	.59
13. Vocabulary Composite	-	-	.56	.56	.64	-.54	.70	.55
14. Word Reading	-	-	.51	.51	.59	-.44	.66	.47
15. Word Meaning	-	-	.45	.46	.56	-.56	.69	.50
16. Comprehension Composite	-	-	.53	.53	.68	-.56	.75	.61
17. Sentence Comprehension	-	-	.56	.56	.65	-.54	.73	.57
18. Passage Comprehension	-	-	.45	.43	.62	-.49	.65	.55
19. Listening Comprehension	-	-	.18*	.18*	.18*	-.15†	.16*	.18†
End of year								
Raw Scores								
1. Total Test	-	-	.56	.56	.75	-.43	.73	.40
2. Vocabulary Composite	-	-	.46	.46	.60	-.43	.73	.27
3. Word Reading	-	-	.39	.38	.48	-.31	.56	.22***

4. Word Meaning	-	-	.43	.43	.57	-.45	.72	.23***
5. Comprehension Composite	-	-	.54	.54	.75	-.39	.65	.41
6. Sentence Comprehension	-	-	.55	.53	.71	-.39	.65	.33
7. Passage Comprehension	-	-	.47	.47	.68	-.36	.57	.41
8. Listening Comprehension	-	-	.25	.21***	.21***	-.09†	.20***	.19*
Standard Scores								
9. Total Test	-	-	.57	.58	.79	-.40	.68	.42
10. Vocabulary Composite	-	-	.57	.56	.72	-.43	.67	.28
11. Comprehension Composite	-	-	.56	.56	.78	-.37	.62	.43
Stanine Scores								
12. Total Test	-	-	.57	.58	.77	-.41	.69	.41
13. Vocabulary Composite	-	-	.56	.55	.72	-.42	.64	.26
14. Word Reading	-	-	.55	.55	.69	-.42	.66	.25
15. Word Meaning	-	-	.49	.49	.66	-.43	.65	.22***
16. Comprehension Composite	-	-	.55	.55	.76	-.40	.64	.43
17. Sentence Comprehension	-	-	.57	.56	.73	-.35	.60	.34
18. Passage Comprehension	-	-	.45	.46	.67	-.34	.55	.42
19. Listening Comprehension	-	-	.21***	.21***	.21***	-.06†	.15*	.25

Note. The GRADE was administered during the end-of-year benchmark assessment. Pair-wise sample sizes for all times of year ≈ 193 . LNF = Letter Naming Fluency; PSF = Phoneme Segmentation Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy. Unless otherwise marked, all correlations are significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant

Table 32

Second Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

[illegible]

12. Total Test	.55	.55	.54	.73	-.44	.67	.55	.45
13. Vocabulary Composite	.52	.51	.51	.66	-.46	.64	.52	.42
14. Word Reading	.48	.45	.46	.62	-.37	.60	.46	.36
15. Word Meaning	.50	.52	.48	.62	-.44	.64	.48	.38
16. Comprehension Composite	.54	.53	.52	.70	-.42	.64	.54	.44
17. Sentence Comprehension	.50	.49	.48	.65	-.44	.63	.50	.40
18. Passage Comprehension	.51	.51	.50	.66	-.40	.62	.50	.40
19. Listening Comprehension	.37	.38	.36	.39	-.27	.41	.25	.12†

Middle of year

Raw Scores

1. Total Test	-	-	-	.76	-.57	.78	.52	.44
2. Vocabulary Composite	-	-	-	.61	-.55	.77	.40	.37
3. Word Reading	-	-	-	.60	-.52	.72	.35	.34
4. Word Meaning	-	-	-	.50	-.49	.68	.34	.27
5. Comprehension Composite	-	-	-	.76	-.54	.72	.52	.42
6. Sentence Comprehension	-	-	-	.71	-.55	.72	.51	.38
7. Passage Comprehension	-	-	-	.71	-.50	.64	.47	.39
8. Listening Comprehension	-	-	-	.42	-.30	.46	.28	.24

Standard Scores								
9. Total Test	-	-	-	.81	-.54	.67	.53	.44
10. Vocabulary Composite	-	-	-	.71	-.57	.69	.44	.41
11. Comprehension Composite	-	-	-	.78	-.51	.64	.53	.42
Stanine Scores								
12. Total Test	-	-	-	.80	-.54	.68	.54	.45
13. Vocabulary Composite	-	-	-	.71	-.56	.68	.45	.41
14. Word Reading	-	-	-	.67	-.52	.64	.40	.36
15. Word Meaning	-	-	-	.67	-.53	.66	.43	.37
16. Comprehension Composite	-	-	-	.77	-.51	.65	.53	.43
17. Sentence Comprehension	-	-	-	.70	-.50	.61	.47	.34
18. Passage Comprehension	-	-	-	.73	-.49	.62	.48	.39
19. Listening Comprehension	-	-	-	.40	-.30	.41	.27	.23
End of year								
Raw Scores								
1. Total Test	-	-	-	.73	-.51	.67	.48	.46
2. Vocabulary Composite	-	-	-	.59	-.58	.75	.36	.37
3. Word Reading	-	-	-	.60	-.56	.71	.37	.38

4. Word Meaning	-	-	-	.48	-.48	.64	.29	.28
5. Comprehension Composite	-	-	-	.73	-.46	.60	.49	.46
6. Sentence Comprehension	-	-	-	.70	-.52	.64	.44	.41
7. Passage Comprehension	-	-	-	.68	-.40	.51	.48	.44
8. Listening Comprehension	-	-	-	.42	-.22	.35	.35	.30
Standard Scores								
9. Total Test	-	-	-	.77	-.45	.56	.53	.48
10. Vocabulary Composite	-	-	-	.69	-.56	.63	.43	.41
11. Comprehension Composite	-	-	-	.76	-.42	.53	.53	.47
Stanine Scores								
12. Total Test	-	-	-	.75	-.45	.55	.51	.48
13. Vocabulary Composite	-	-	-	.68	-.54	.61	.43	.41
14. Word Reading	-	-	-	.66	-.52	.61	.41	.41
15. Word Meaning	-	-	-	.64	-.53	.61	.40	.35
16. Comprehension Composite	-	-	-	.74	-.42	.52	.52	.46
17. Sentence Comprehension	-	-	-	.69	-.44	.51	.42	.39
18. Passage Comprehension	-	-	-	.69	-.41	.50	.49	.45
19. Listening Comprehension	-	-	-	.40	-.23	.30	.35	.27

Note. The GRADE was administered during the end-of-year benchmark assessment. Pair-wise sample sizes for all times of year ≈ 215 . CLS = Correct Letter Sounds; WWR = Whole Words Read; TW = Total Words Attempted; DORF = DIBELS Oral Reading Fluency Words Correct; DORFE = DIBELS Oral Reading Fluency Errors; DORFA = DIBELS Oral Reading Fluency Accuracy; RQR = Retell Quality of Response. Unless otherwise marked, all correlations significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant.

Table 33

Third Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

Subtests and Composites	ORF	ORFE	ORFA	RTF	RQR	DZA
Beginning of year						
Raw Scores						
1. Total Score	.66	-.57	.68	.48	.39	.67
2. Vocabulary Composite	.59	-.57	.69	.41	.33	.59
3. Word Reading	.52	-.60	.73	.34	.24	.51
4. Vocabulary	.56	-.50	.61	.40	.32	.56
5. Comprehension Composite	.66	-.51	.60	.50	.38	.67
6. Sentence Comprehension	.43	-.48	.59	.28	.30	.48
7. Passage Comprehension	.68	-.44	.51	.54	.36	.67
8. Listening Comprehension	.24	-.06†	.10†	.28	.23	.22**
Standard Scores						
9. Total Score	.68	-.51	.58	.53	.40	.69
10. Vocabulary Composite	.65	-.51	.60	.49	.37	.66

11. Comprehension Composite	.65	-.46	.53	.51	.38	.67
Stanines						
12. Total Score	.68	-.50	.57	.52	.40	.70
13. Vocabulary Composite	.66	-.52	.60	.48	.37	.65
14. Word Reading	.58	-.51	.57	.37	.23	.61
15. Vocabulary	.63	-.48	.59	.48	.37	.63
16. Comprehension Composite	.65	-.45	.51	.50	.39	.66
17. Sentence Comprehension	.51	-.44	.50	.40	.38	.56
18. Passage Comprehension	.66	-.41	.47	.55	.37	.65
19. Listening Comprehension	.23***	-.07†	.10†	.26	.21***	.19*
Middle of year						
Raw Scores						
1. Total Score	.67	-.54	.71	.56	.48	.65
2. Vocabulary Composite	.60	-.53	.71	.46	.40	.60
3. Word Reading	.52	-.43	.65	.36	.31	.49
4. Vocabulary	.57	-.51	.65	.45	.38	.58
5. Comprehension Composite	.68	-.49	.64	.59	.50	.63
6. Sentence Comprehension	.44	-.38	.55	.37	.36	.47

7. Passage Comprehension	.69	-.47	.56	.62	.49	.62
8. Listening Comprehension	.25	.09†	.13†	.23	.23	.24
Standard Scores						
9. Total Score	.69	-.51	.63	.62	.51	.67
10. Vocabulary Composite	.65	-.51	.64	.55	.47	.66
11. Comprehension Composite	.66	-.47	.58	.61	.49	.62
Stanines						
12. Total Score	.68	-.51	.63	.62	.52	.67
13. Vocabulary Composite	.65	-.51	.65	.55	.47	.67
14. Word Reading	.58	-.40	.54	.37	.33	.58
15. Vocabulary	.64	-.52	.63	.54	.45	.63
16. Comprehension Composite	.66	-.46	.57	.60	.50	.63
17. Sentence Comprehension	.53	-.41	.52	.51	.47	.53
18. Passage Comprehension	.67	-.46	.53	.63	.49	.60
19. Listening Comprehension	.22***	-.10†	.13†	.24	.26	.20**
End of year						
Raw Scores						
1. Total Score	.66	-.48	.59	.53	.55	.67

2. Vocabulary Composite	.58	-.46	.61	.44	.47	.61
3. Word Reading	.47	-.36	.58	.35	.40	.54
4. Vocabulary	.57	-.46	.57	.43	.45	.56
5. Comprehension Composite	.66	-.45	.52	.57	.58	.65
6. Sentence Comprehension	.44	-.40	.51	.38	.42	.44
7. Passage Comprehension	.68	-.44	.46	.58	.57	.66
8. Listening Comprehension	.23***	-.22***	.18*	.27	.26	.29
Standard Scores						
9. Total Score	.67	-.45	.52	.56	.55	.67
10. Vocabulary Composite	.62	-.42	.53	.51	.50	.66
11. Comprehension Composite	.65	-.43	.48	.57	.55	.63
Stanines						
12. Total Score	.67	-.46	.52	.57	.56	.67
13. Vocabulary Composite	.63	-.42	.53	.52	.51	.66
14. Word Reading	.52	-.25	.39	.36	.37	.61
15. Vocabulary	.63	-.47	.55	.51	.49	.62
16. Comprehension Composite	.65	-.43	.47	.55	.55	.63
17. Sentence Comprehension	.52	-.39	.45	.49	.51	.48

18. Passage Comprehension	.66	-.42	.44	.58	.55	.64
19. Listening Comprehension	.20***	-.16*	.14†	.27	.26	.26

Note. Pair-wise sample sizes for all times of year ≈ 195 . ORF = DIBELS Oral Reading Fluency Words Correct; ORFE = DIBELS Oral Reading Fluency Errors; ORFA = DIBELS Oral Reading Fluency Accuracy; RTF = Retell; RQR = Retell Quality of Response; DZA = Daze Adjusted Score. Unless otherwise marked, all correlations are significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant.

Table 34

Fourth Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

Subtests and Composites	ORF	ORFE	ORFA	RTF	RQR	DZA
Beginning of year						
Raw Scores						
1. Total Score	.77	-.48	.62	.56	.37	.68
2. Vocabulary	.73	-.50	.63	.55	.42	.64
3. Comprehension Composite	.74	-.45	.60	.52	.32	.67
4. Sentence Comprehension	.75	-.49	.66	.52	.37	.66
5. Passage Comprehension	.65	-.38	.49	.47	.26	.60
Standard Scores						
6. Total Score	.76	-.46	.58	.55	.36	.69
7. Vocabulary	.72	-.48	.59	.55	.41	.64
8. Comprehension Composite	.74	-.44	.56	.52	.30	.68
Stanines						
9. Total Score	.74	-.46	.57	.54	.35	.68

10. Vocabulary	.72	-.48	.60	.56	.41	.63
11. Comprehension Composite	.74	-.43	.54	.52	.32	.68
12. Sentence Comprehension	.77	-.46	.59	.52	.35	.66
13. Passage Comprehension	.64	-.38	.47	.46	.25	.59
Middle of year						
Raw Scores						
1. Total Score	.77	-.52	.62	.63	.56	.63
2. Vocabulary	.75	-.55	.64	.60	.54	.62
3. Comprehension Composite	.74	-.47	.57	.61	.55	.61
4. Sentence Comprehension	.77	-.49	.62	.59	.55	.61
5. Passage Comprehension	.64	-.41	.48	.57	.50	.54
Standard Scores						
6. Total Score	.76	-.50	.58	.64	.54	.65
7. Vocabulary	.73	-.53	.61	.60	.52	.62
8. Comprehension Composite	.74	-.44	.54	.62	.54	.63
Stanines						
9. Total Score	.74	-.49	.56	.62	.52	.62
10. Vocabulary	.73	-.53	.61	.61	.53	.62

11. Comprehension Composite	.73	-.43	.52	.62	.52	.64
12. Sentence Comprehension	.77	-.46	.56	.62	.53	.63
13. Passage Comprehension	.62	-.40	.46	.55	.49	.54
End of year						
Raw Scores						
1. Total Score	.74	-.46	.54	.62	.61	.68
2. Vocabulary	.71	-.47	.53	.59	.60	.66
3. Comprehension Composite	.73	-.44	.52	.60	.58	.66
4. Sentence Comprehension	.76	-.50	.60	.54	.55	.68
5. Passage Comprehension	.62	-.36	.42	.56	.54	.56
Standard Scores						
6. Total Score	.72	-.45	.50	.63	.59	.67
7. Vocabulary	.69	-.46	.51	.60	.59	.65
8. Comprehension Composite	.72	-.43	.49	.61	.56	.65
Stanines						
9. Total Score	.71	-.44	.49	.61	.59	.64
10. Vocabulary	.69	-.46	.51	.60	.59	.65
11. Comprehension Composite	.71	-.40	.46	.62	.57	.66

12. Sentence Comprehension	.75	-.46	.53	.60	.55	.66
13. Passage Comprehension	.61	-.35	.40	.57	.52	.56

Note. Pair-wise sample sizes for all times of year ≈ 180 . ORF = DIBELS Oral Reading Fluency Words Correct; ORFE = DIBELS Oral Reading Fluency Errors; ORFA = DIBELS Oral Reading Fluency Accuracy; RTF = Retell; RQR = Retell Quality of Response; DZA = Daze Adjusted Score. Unless otherwise marked, all correlations significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant.

Table 35

Fifth Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

Subtests and Composites	ORF	ORFE	ORFA	RTF	RQR	DZA
Beginning of year						
Raw Scores						
1. Total Score	.69	-.49	.53	.61	.44	.61
2. Vocabulary	.68	-.54	.51	.57	.45	.60
3. Comprehension Composite	.65	-.45	.51	.59	.41	.58
4. Sentence Comprehension	.66	-.52	.54	.54	.44	.58
5. Passage Comprehension	.58	-.37	.42	.56	.35	.52
Standard Scores						
6. Total Score	.67	-.48	.50	.63	.45	.59
7. Vocabulary	.67	-.54	.51	.57	.45	.59
8. Comprehension Composite	.62	-.43	.47	.64	.41	.56
Stanines						
9. Total Score	.68	-.48	.52	.62	.44	.60

10. Vocabulary	.66	-.52	.49	.58	.46	.60
11. Comprehension Composite	.63	-.42	.47	.62	.40	.56
12. Sentence Comprehension	.63	-.48	.49	.57	.41	.54
13. Passage Comprehension	.55	-.36	.41	.58	.35	.52
Middle of year						
Raw Scores						
1. Total Score	.65	-.49	.49	.63	.50	.65
2. Vocabulary	.64	-.49	.47	.61	.48	.64
3. Comprehension Composite	.61	-.48	.48	.60	.49	.62
4. Sentence Comprehension	.63	-.49	.50	.56	.47	.60
5. Passage Comprehension	.54	-.42	.41	.57	.45	.58
Standard Scores						
6. Total Score	.63	-.47	.46	.64	.49	.64
7. Vocabulary	.64	-.50	.48	.61	.48	.64
8. Comprehension Composite	.59	-.45	.45	.63	.47	.60
Stanines						
9. Total Score	.65	-.47	.47	.64	.49	.65
10. Vocabulary	.63	-.48	.45	.61	.48	.63

11. Comprehension Composite	.59	-.45	.44	.62	.46	.60
12. Sentence Comprehension	.60	-.48	.47	.61	.48	.57
13. Passage Comprehension	.52	-.41	.40	.59	.45	.57
End of year						
Raw Scores						
1. Total Score	.65	-.48	.49	.65	.62	.66
2. Vocabulary	.65	-.48	.48	.61	.60	.67
3. Comprehension Composite	.62	-.47	.49	.64	.60	.62
4. Sentence Comprehension	.64	-.51	.54	.58	.56	.61
5. Passage Comprehension	.55	-.39	.41	.61	.57	.56
Standard Scores						
6. Total Score	.63	-.46	.46	.66	.60	.66
7. Vocabulary	.65	-.48	.49	.61	.59	.67
8. Comprehension Composite	.59	-.43	.44	.66	.59	.62
Stanines						
9. Total Score	.65	-.47	.48	.64	.61	.66
10. Vocabulary	.64	-.46	.46	.61	.61	.66
11. Comprehension Composite	.59	-.43	.44	.65	.59	.60

12. Sentence Comprehension	.60	-.45	.47	.58	.52	.59
13. Passage Comprehension	.52	-.39	.40	.62	.56	.56

Note. Pair-wise sample sizes for all times of year ≈ 195 . ORF = DIBELS Oral Reading Fluency Words Correct; ORFE = DIBELS Oral Reading Fluency Errors; ORFA = DIBELS Oral Reading Fluency Accuracy; RTF = Retell; RQR = Retell Quality of Response; DZA = Daze Adjusted Score. Unless otherwise marked, all correlations significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant.

Table 36

Sixth Grade GRADE Subtest and Composite Score Correlations with All DIBELS Next Measures

Subtests and Composites	ORF	ORFE	ORFA	RTF	RQR	DZA
Beginning of year						
Raw Scores						
1. Total Score	.64	-.43	.55	.55	.50	.61
2. Vocabulary	.63	-.41	.51	.58	.44	.61
3. Comprehension Composite	.58	-.40	.53	.48	.50	.55
4. Sentence Comprehension	.56	-.45	.56	.50	.43	.51
5. Passage Comprehension	.53	-.32	.45	.41	.48	.51
Standard Scores						
6. Total Score	.62	-.41	.51	.53	.51	.59
7. Vocabulary	.62	-.40	.49	.57	.44	.60
8. Comprehension Composite	.55	-.38	.47	.42	.49	.52
Stanines						
9. Total Score	.64	-.42	.53	.58	.52	.59

10. Vocabulary	.64	-.41	.51	.58	.42	.62
11. Comprehension Composite	.57	-.38	.48	.46	.51	.54
12. Sentence Comprehension	.55	-.42	.49	.50	.39	.50
13. Passage Comprehension	.49	-.27***	.39	.37	.48	.46
Middle of year						
Raw Scores						
1. Total Score	.59	-.33	.47	.59	.59	.58
2. Vocabulary	.58	-.33	.44	.56	.50	.59
3. Comprehension Composite	.55	-.30	.44***	.56	.58	.51
4. Sentence Comprehension	.53	-.38	.49	.51	.50	.47
5. Passage Comprehension	.49	-.21	.36*	.52	.55	.48
Standard Scores						
6. Total Score	.58	-.29	.42***	.59	.58	.58
7. Vocabulary	.57	-.32	.42	.55	.50	.59
8. Comprehension Composite	.53	-.27	.39***	.53	.57	.49
Stanines						
9. Total Score	.58	-.31	.45	.62	.59	.57
10. Vocabulary	.60	-.36	.47	.56	.48	.60

11. Comprehension Composite	.55	-.27	.39***	.57	.60	.52
12. Sentence Comprehension	.54	-.36	.44	.51	.47	.46
13. Passage Comprehension	.46	-.18	.31†	.47	.51	.44
End of year						
Raw Scores						
1. Total Score	.61	-.48	.55	.56	.60	.64
2. Vocabulary	.59	-.47	.52	.57	.57	.60
3. Comprehension Composite	.57	-.44	.53	.51	.56	.62
4. Sentence Comprehension	.53	-.46	.54	.53	.53	.52
5. Passage Comprehension	.53	-.37	.46	.44	.52	.60
Standard Scores						
6. Total Score	.59	-.45	.51	.55	.60	.61
7. Vocabulary	.57	-.46	.50	.56	.56	.59
8. Comprehension Composite	.56	-.41	.46	.46	.55	.58
Stanines						
9. Total Score	.60	-.45	.51	.60	.60	.63
10. Vocabulary	.60	-.47	.53	.58	.53	.60
11. Comprehension Composite	.57	-.41	.47	.48	.56	.60

12. Sentence Comprehension	.51	-.45	.48	.48	.50	.47
13. Passage Comprehension	.51	-.33	.40	.41	.51	.55

Note. Pair-wise sample sizes for all times of year ≈ 105 . ORF = DIBELS Oral Reading Fluency Words Correct; ORFE = DIBELS Oral Reading Fluency Errors; ORFA = DIBELS Oral Reading Fluency Accuracy; RTF = Retell; RQR = Retell Quality of Response; DZA = Daze Adjusted Score. Unless otherwise marked, all correlations significant at $p < .001$.

*** $p < .005$

** $p < .01$

* $p < .05$

† = Not Significant.

Correlations between individual DIBELS Next measures at each grade level and DIBELS Composite Score (DCS) at all three benchmark periods are presented in Tables 37 through 43. In kindergarten, LNF and NWF correlate the strongest with the DCS, with values ranging from .60 for beginning of year LNF with middle of year DCS to .88 for end of year NWF - CLS with end of year DCS. In first through sixth grade, DORF - Words Correct correlates the strongest with the DCS at each time of year, with significant correlations also coming from the Daze Adjusted Score in grades three through six.

Table 37

BGS Kindergarten DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
First Sound Fluency	.86 (476)	.57 (467)	.43 (457)
Letter Naming Fluency	.91 (476)	.60 (467)	.49 (457)
Middle of Year			
First Sound Fluency	.47 (467)	.76 (476)	.47 (464)
Letter Naming Fluency	.60 (467)	.81 (476)	.60 (464)
Phoneme-Segmentation Fluency	.45 (467)	.80 (476)	.47 (464)
NWF: Correct Letter Sounds	.59 (467)	.79 (476)	.65 (464)
NWF: Whole Words Read	.42 (465)	.49 (474)	.52 (462)
NWF: Total Words Attempted	.54 (454)	.69 (463)	.55 (451)
End of Year			
Letter Naming Fluency	.46 (458)	.67 (466)	.83 (477)
Phoneme-Segmentation Fluency	.26 (458)	.45 (466)	.72 (477)

NWF: Correct Letter Sounds	.51 (457)	.60 (464)	.88 (477)
NWF: Whole Words Read	.42 (457)	.51 (464)	.74 (477)
NWF: Total Words Attempted	.53 (447)	.59 (454)	.85 (466)

Note. N = 484. DIBELS Composite Score is comprised of LNF and PSF at beginning of year; FSF, LNF, PSF, and NWF: CLS at middle of year; LNF, PSF, and NWF:CLS at end of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 38

BGS First Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
Letter Naming Fluency	.82 (460)	.65 (445)	.65 (439)
Phoneme-Segmentation Fluency	.63 (460)	.25 (445)	.26 (439)
NWF: Correct Letter Sounds	.88 (460)	.82 (445)	.71 (439)
NWF: Whole Words Read	.75 (460)	.79 (445)	.66 (439)
NWF: Total Words Attempted	.87 (447)	.82 (433)	.69 (428)
Middle of Year			
NWF: Correct Letter Sounds	.75 (451)	.92 (450)	.78 (445)
NWF: Whole Words Read	.71 (450)	.90 (450)	.78 (444)
NWF: Total Words Attempted	.76 (442)	.92 (441)	.78 (436)
DORF: Words Correct	.76 (451)	.94 (450)	.83 (445)
DORF: Errors	-.42 (451)	-.68 (450)	-.54 (444)
DORF: Accuracy	.63 (446)	.83 (450)	.81 (441)

Retell	.61 (264)	.73 (264)	.67 (262)
Retell Quality of Response	.45 (138)	.51 (140)	.44 (137)
End of Year			
NWF: Correct Letter Sounds	.72 (441)	.82 (441)	.86 (459)
NWF: Whole Words Read	.67 (441)	.81 (441)	.87 (459)
NWF: Total Words Attempted	.71 (437)	.81 (437)	.84 (455)
DORF: Words Correct	.71 (439)	.89 (440)	.93 (459)
DORF: Errors	-.43 (439)	-.49 (440)	-.66 (459)
DORF: Accuracy	.54 (439)	.64 (440)	.84 (459)
Retell	.32 (358)	.38 (359)	.39 (373)
Retell Quality of Response	.23 (356)	.30 (357)	.33 (371)

Note. N = 475. DIBELS Composite Score is comprised of LNF, PSF, and NWF: CLS at beginning of year; NWF: CLS, NWF: WWR, DORF: Words Correct, and a scaled score based on DORF: Accuracy at middle of year; NWF: WWR, DORF: Words Correct, and a scaled score based on DORF: Accuracy at end of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 39

BGS Second Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
NWF: Correct Letter Sounds	.86 (558)	.69 (548)	.65 (539)
NWF: Whole Words Read	.86 (558)	.65 (548)	.62 (539)
NWF: Total Words Attempted	.84 (551)	.68 (541)	.64 (532)

DORF: Words Correct	.94 (558)	.85 (548)	.81 (539)
DORF: Errors	-.62 (558)	-.48 (547)	-.41 (538)
DORF: Accuracy	.82 (558)	.75 (546)	.71 (537)
Retell	.58 (519)	.63 (509)	.62 (501)
Retell Quality of Response	.43 (473)	.49 (463)	.49 (455)
Middle of Year			
DORF: Words Correct	.89 (546)	.93 (564)	.87 (554)
DORF: Errors	-.61 (546)	-.72 (564)	-.61 (554)
DORF: Accuracy	.70 (546)	.79 (564)	.75 (554)
Retell	.55 (530)	.81 (547)	.70 (537)
Retell Quality of Response	.40 (505)	.62 (520)	.56 (510)
End of Year			
DORF: Words Correct	.85 (538)	.88 (555)	.92 (565)
DORF: Errors	-.55 (538)	-.57 (555)	-.68 (565)
DORF: Accuracy	.59 (538)	.64 (555)	.75 (565)
Retell	.50 (519)	.65 (533)	.82 (541)
Retell Quality of Response	.45 (519)	.58 (533)	.67 (541)

Note. N = 578. DIBELS Composite Score is comprised of NWF: WWR, DORF: Words Correct, and a scaled score based on DORF: Accuracy at beginning of year; DORF: Retell, DORF: Words Correct, and a scaled score based on DORF: Accuracy at middle of year and end of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 40

BGS Third Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
DORF: Words Correct	.93 (501)	.88 (491)	.86 (481)
DORF: Errors	-.63 (501)	-.55 (491)	-.51 (481)
DORF: Accuracy	.74 (501)	.71 (491)	.70 (481)
Retell	.77 (478)	.67 (469)	.64 (460)
Retell Quality of Response	.57 (465)	.49 (456)	.46 (450)
DIBELS Maze (Daze)	.80 (501)	.76 (490)	.70 (480)
Daze Errors	-.42 (501)	-.40 (490)	-.45 (480)
Daze Adjusted Score	.84 (501)	.79 (490)	.74 (480)
Middle of Year			
DORF: Words Correct	.88 (491)	.90 (501)	.86 (487)
DORF: Errors	-.58 (491)	-.63 (501)	-.55 (487)
DORF: Accuracy	.65 (491)	.72 (501)	.69 (487)
Retell	.63 (479)	.78 (489)	.65 (476)
Retell Quality of Response	.53 (468)	.66 (478)	.55 (467)
DIBELS Maze (Daze)	.72 (490)	.80 (501)	.69 (486)
Daze Errors	-.37 (490)	-.38 (501)	-.35 (486)
Daze Adjusted Score	.75 (490)	.83 (501)	.72 (486)
End of Year			
DORF: Words Correct	.87 (482)	.87 (489)	.91 (495)
DORF: Errors	-.54 (482)	-.54 (489)	-.63 (495)

DORF: Accuracy	.61 (482)	.64 (489)	.73 (495)
Retell	.58 (472)	.62 (478)	.75 (486)
Retell Quality of Response	.51 (471)	.54 (477)	.61 (485)
DIBELS Maze (Daze)	.76 (480)	.79 (488)	.84 (495)
Daze Errors	-.36 (480)	-.35 (488)	-.43 (495)
Daze Adjusted Score	.79 (480)	.80 (488)	.87 (495)

Note. N = 513. DIBELS Composite Score is comprised of DORF: Retell, DORF: Words Correct, a scaled score based on DORF: Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 41

BGS Fourth Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
DORF: Words Correct	.93 (584)	.90 (574)	.86 (573)
DORF: Errors	-.59 (584)	-.53 (574)	-.50 (573)
DORF: Accuracy	.74 (584)	.72 (574)	.71 (573)
Retell	.75 (572)	.62 (562)	.62 (562)
Retell Quality of Response	.53 (561)	.45 (551)	.42 (551)
DIBELS Maze (Daze)	.82 (584)	.75 (572)	.74 (570)
Daze Errors	-.29 (584)	-.29 (572)	-.25 (570)
Daze Adjusted Score	.84 (584)	.76 (572)	.76 (570)
Middle of Year			

DORF: Words Correct	.90 (578)	.93 (585)	.87 (580)
DORF: Errors	-.57 (578)	-.62 (585)	-.53 (580)
DORF: Accuracy	.66 (578)	.72 (585)	.68 (580)
Retell	.66 (574)	.77 (581)	.68 (577)
Retell Quality of Response	.55 (507)	.63 (512)	.53 (510)
DIBELS Maze (Daze)	.75 (571)	.82 (585)	.73 (574)
Daze Errors	-.33 (571)	-.37 (585)	-.34 (574)
Daze Adjusted Score	.77 (571)	.84 (585)	.75 (574)
End of Year			
DORF: Words Correct	.87 (573)	.88 (577)	.91 (585)
DORF: Errors	-.29 (573)	-.27 (577)	-.38 (585)
DORF: Accuracy	.55 (573)	.55 (577)	.63 (585)
Retell	.59 (567)	.59 (570)	.76 (579)
Retell Quality of Response	.54 (566)	.57 (569)	.61 (578)
DIBELS Maze (Daze)	.78 (571)	.79 (575)	.85 (585)
Daze Errors	-.31 (571)	-.35 (575)	-.35 (585)
Daze Adjusted Score	.79 (571)	.81 (575)	.86 (585)

Note. N = 599. DIBELS Composite Score is comprised of DORF: Retell, DORF: Words Correct, a scaled score based on DORF: Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 42

BGS Fifth Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
DORF: Words Correct	.91 (541)	.89 (530)	.85 (516)
DORF: Errors	-.69 (541)	-.62 (530)	-.58 (516)
DORF: Accuracy	.73 (541)	.69 (530)	.66 (516)
Retell	.72 (538)	.58 (526)	.58 (512)
Retell Quality of Response	.60 (534)	.45 (522)	.44 (509)
DIBELS Maze (Daze)	.82 (541)	.72 (529)	.67 (514)
Daze Errors	-.55 (541)	-.48 (529)	-.50 (514)
Daze Adjusted Score	.84 (541)	.74 (529)	.69 (514)
Middle of Year			
DORF: Words Correct	.85 (533)	.92 (535)	.87 (520)
DORF: Errors	-.63 (533)	-.66 (535)	-.60 (520)
DORF: Accuracy	.66 (533)	.69 (535)	.64 (520)
Retell	.65 (532)	.80 (535)	.67 (519)
Retell Quality of Response	.53 (528)	.62 (531)	.54 (517)
DIBELS Maze (Daze)	.75 (528)	.83 (535)	.74 (514)
Daze Errors	-.42 (528)	-.41 (535)	-.42 (514)
Daze Adjusted Score	.79 (528)	.86 (535)	.77 (514)
End of Year			
DORF: Words Correct	.85 (518)	.88 (519)	.90 (532)
DORF: Errors	-.58 (518)	-.56 (519)	-.64 (532)

DORF: Accuracy	.66 (518)	.66 (519)	.71 (532)
Retell	.52 (511)	.57 (513)	.72 (528)
Retell Quality of Response	.51 (510)	.53 (512)	.64 (527)
DIBELS Maze (Daze)	.76 (516)	.79 (517)	.85 (532)
Daze Errors	-.26 (516)	-.27 (517)	-.37 (532)
Daze Adjusted Score	.76 (516)	.79 (517)	.87 (532)

Note. N = 558. DIBELS Composite Score is comprised of DORF: Retell, DORF: Words Correct, a scaled score based on DORF: Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 43

BGS Sixth Grade DIBELS Composite Score Correlations with All DIBELS Next Measures

Measure by Time of Year	DCS by Benchmark Assessment		
	Beg.	Mid.	End
Beginning of Year			
DORF: Words Correct	.91 (534)	.87 (510)	.86 (513)
DORF: Errors	-.59 (534)	-.52 (510)	-.46 (513)
DORF: Accuracy	.73 (534)	.66 (510)	.65 (513)
Retell	.73 (531)	.61 (506)	.62 (510)
Retell Quality of Response	.50 (529)	.45 (504)	.45 (508)
DIBELS Maze (Daze)	.83 (534)	.78 (510)	.76 (513)
Daze Errors	-.27 (534)	-.28 (510)	-.25 (513)
Daze Adjusted Score	.84 (534)	.78 (510)	.77 (513)
Middle of Year			

DORF: Words Correct	.88 (517)	.92 (519)	.87 (516)
DORF: Errors	-.55 (517)	-.64 (519)	-.50 (516)
DORF: Accuracy	.66 (517)	.68 (519)	.62 (516)
Retell	.70 (515)	.80 (517)	.72 (515)
Retell Quality of Response	.59 (511)	.66 (513)	.60 (511)
DIBELS Maze (Daze)	.75 (510)	.86 (519)	.76 (508)
Daze Errors	-.46 (510)	-.50 (519)	-.42 (508)
Daze Adjusted Score	.77 (510)	.87 (519)	.77 (508)
End of Year			
DORF: Words Correct	.88 (517)	.88 (512)	.91 (526)
DORF: Errors	-.55 (517)	-.57 (512)	-.57 (526)
DORF: Accuracy	.67 (517)	.65 (512)	.68 (526)
Retell	.61 (513)	.62 (508)	.77 (523)
Retell Quality of Response	.53 (511)	.57 (506)	.66 (521)
DIBELS Maze (Daze)	.75 (514)	.78 (509)	.85 (526)
Daze Errors	-.43 (514)	-.43 (509)	-.38 (526)
Daze Adjusted Score	.76 (514)	.79 (509)	.86 (526)

Note. N = 551. DIBELS Composite Score is comprised of DORF: Retell, DORF: Words Correct, a scaled score based on DORF: Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures were not given a composite score. All correlations significant at $p < .001$.

Table 44

BGS Kindergarten DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.50 (166)	.48 (168)	.37 (166)
2. Phonological Awareness Composite	.48 (177)	.42 (180)	.31 (178)
3. Sound Matching	.41 (177)	.40 (180)	.39 (178)
4. Rhyming	.43 (179)	.32 (182)	.14* (180)
5. Early Literacy Skills Composite	.32 (168)	.27 (170)	.15* (168)
6. Print Awareness	.31 (177)	.24** (180)	.16* (178)
7. Letter Recognition	.13† (168)	.09† (170)	.06† (168)
8. Same & Different Words	.31 (179)	.30 (182)	.12† (180)
9. Phoneme-Grapheme Correspondence	.35 (177)	.43 (180)	.34 (178)
10. Listening Comprehension	.27 (179)	.36 (182)	.36 (180)
Standard Scores			
11. Total Test	.57 (166)	.51 (168)	.43 (166)
Stanine Scores			
12. Total Test	.57 (166)	.50 (168)	.42 (166)
13. Phonological Awareness Composite	.49 (177)	.42 (180)	.36 (178)
14. Early Literacy Skills Composite	.43 (168)	.34 (170)	.25 (168)
15. Phoneme-Grapheme Correspondence	.41 (177)	.49 (180)	.39 (178)
16. Listening Comprehension	.34 (179)	.41 (182)	.40 (180)

Note. N = 182. Kindergarten DIBELS Composite Score is comprised of LNF and PSF at beginning of year; FSF, LNF, PSF, and NWF-CLS at middle of year; LNF, PSF, and NWF-CLS at end of year.

Students who did not have scores in all contributing measures for DIBELS were not given a composite

score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score.

Unless marked, all correlations significant at $p < .001$. Significance codes:

* $p < .05$

** $p < .01$

† Not Significant.

Table 45

BGS First Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.55 (193)	.71 (195)	.77 (193)
2. Vocabulary Composite	.46 (196)	.57 (198)	.63 (196)
3. Word Reading	.39 (199)	.47 (200)	.50 (199)
4. Word Meaning	.41 (196)	.53 (198)	.60 (196)
5. Comprehension Composite	.55 (193)	.71 (195)	.71 (193)
6. Sentence Comprehension	.52 (199)	.68 (200)	.69 (199)
7. Passage Comprehension	.48 (193)	.63 (195)	.63 (193)
8. Listening Comprehension	.23 (200)	.22* (201)	.22** (200)
Standard Scores			
9. Total Test	.58 (193)	.76 (195)	.75 (193)
10. Vocabulary Composite	.56 (196)	.70 (198)	.71 (196)
11. Comprehension Composite	.56 (193)	.74 (195)	.72 (193)
Stanine Scores			
12. Total Test	.57 (193)	.74 (195)	.74 (193)
13. Vocabulary Composite	.56 (196)	.70 (198)	.70 (196)

14. Word Reading	.51 (199)	.65 (200)	.69 (199)
15. Word Meaning	.47 (196)	.63 (198)	.65 (196)
16. Comprehension Composite	.55 (193)	.72 (195)	.72 (193)
17. Sentence Comprehension	.53 (199)	.71 (200)	.70 (199)
18. Passage Comprehension	.46 (193)	.62 (195)	.61 (193)
19. Listening Comprehension	.22 (200)	.21** (201)	.21** (200)

Note. N = 203. First-grade DIBELS Composite Score is comprised of LNF, PSF, and NWF-CLS at beginning of year; NWF-CLS, NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at middle of year; NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at end of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. Unless marked, all correlations significant at $p < .001$. Significance codes:

* $p < .05$

** $p < .01$

Table 46

BGS Second Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.75 (214)	.80 (219)	.75 (217)
2. Vocabulary Composite	.59 (216)	.66 (221)	.65 (219)
3. Word Reading	.59 (222)	.65 (227)	.66 (225)
4. Word Meaning	.51 (217)	.56 (222)	.53 (220)
5. Comprehension Composite	.74 (215)	.79 (220)	.73 (218)
6. Sentence Comprehension	.69 (217)	.77 (222)	.72 (220)

7. Passage Comprehension	.70 (216)	.72 (221)	.67 (219)
8. Listening Comprehension	.45 (223)	.45 (228)	.43 (226)
Standard Scores			
9. Total Test	.75 (214)	.79 (219)	.75 (217)
10. Vocabulary Composite	.70 (216)	.73 (221)	.70 (219)
11. Comprehension Composite	.74 (215)	.77 (220)	.74 (218)
Stanine Scores			
12. Total Test	.75 (214)	.79 (219)	.74 (217)
13. Vocabulary Composite	.70 (216)	.73 (221)	.69 (219)
14. Word Reading	.64 (222)	.68 (227)	.68 (225)
15. Word Meaning	.68 (217)	.70 (222)	.66 (220)
16. Comprehension Composite	.72 (215)	.77 (220)	.72 (218)
17. Sentence Comprehension	.68 (217)	.71 (222)	.66 (220)
18. Passage Comprehension	.69 (216)	.72 (221)	.68 (219)
19. Listening Comprehension	.45 (223)	.43 (228)	.41 (226)

Note. N = 228. Second-grade DIBELS Composite Score is comprised of NWF-WWR, DORF Words Correct, and a scaled score based on DORF Accuracy at beginning of year; Retell, DORF Words Correct, and a scaled score based on DORF Accuracy at middle of year and end of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. All correlations significant at $p < .001$.

Table 47

BGS Third Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.73 (184)	.78 (187)	.75 (187)
2. Vocabulary Composite	.66 (186)	.70 (189)	.67 (189)
3. Word Reading	.59 (188)	.59 (191)	.55 (191)
4. Word Meaning	.64 (188)	.62 (191)	.57 (191)
5. Vocabulary	.62 (189)	.67 (192)	.63 (191)
6. Comprehension Composite	.72 (184)	.77 (187)	.75 (187)
7. Sentence Comprehension	.51 (186)	.55 (189)	.52 (189)
8. Passage Comprehension	.72 (189)	.75 (192)	.75 (191)
9. Listening Comprehension	.26 (190)	.27 (193)	.31 (193)
Standard Scores			
10. Total Test	.74 (184)	.79 (187)	.76 (187)
11. Vocabulary Composite	.71 (186)	.75 (189)	.71 (189)
12. Comprehension Composite	.70 (184)	.75 (187)	.73 (187)
Stanine Scores			
13. Total Test	.74 (184)	.79 (187)	.76 (187)
14. Vocabulary Composite	.71 (186)	.75 (189)	.72 (189)
15. Vocabulary	.69 (189)	.73 (192)	.70 (191)
16. Comprehension Composite	.70 (184)	.75 (187)	.72 (187)
17. Sentence Comprehension	.59 (186)	.64 (189)	.60 (189)
18. Passage Comprehension	.70 (189)	.74 (192)	.73 (191)
19. Listening Comprehension	.24 (190)	.25 (193)	.27 (193)

Note. N = 193. Third-grade DIBELS Composite Score is comprised of Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. All correlations significant at $p < .001$

Table 48

BGS Fourth Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.80 (183)	.80 (185)	.80 (186)
2. Vocabulary Composite	.77 (189)	.78 (190)	.76 (191)
3. Comprehension Composite	.77 (183)	.76 (185)	.77 (186)
4. Sentence Comprehension	.78 (187)	.77 (188)	.78 (189)
5. Passage Comprehension	.67 (183)	.67 (185)	.68 (186)
Standard Scores			
6. Total Test	.78 (183)	.79 (185)	.79 (186)
7. Vocabulary Composite	.76 (189)	.76 (190)	.75 (191)
8. Comprehension Composite	.75 (183)	.76 (185)	.77 (186)
Stanine Scores			
9. Total Test	.77 (183)	.77 (185)	.76 (186)
10. Vocabulary Composite	.76 (189)	.77 (190)	.75 (191)
11. Comprehension Composite	.76 (183)	.75 (185)	.76 (186)
12. Sentence Comprehension	.77 (187)	.76 (188)	.77 (189)

13. Passage Comprehension	.66 (183)	.65 (185)	.67 (186)

Note. N = 193. Fourth-grade DIBELS Composite Score is comprised of Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. All correlations significant at $p < .001$

Table 49

BGS Fifth Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.76 (194)	.76 (194)	.77 (193)
2. Vocabulary Composite	.73 (198)	.73 (199)	.75 (198)
3. Comprehension Composite	.73 (194)	.72 (194)	.74 (193)
4. Sentence Comprehension	.73 (197)	.70 (197)	.73 (196)
5. Passage Comprehension	.65 (194)	.66 (194)	.67 (193)
Standard Scores			
6. Total Test	.75 (194)	.75 (194)	.77 (193)
7. Vocabulary Composite	.72 (198)	.73 (199)	.75 (198)
8. Comprehension Composite	.71 (194)	.71 (194)	.73 (193)
Stanine Scores			
9. Total Test	.76 (194)	.76 (194)	.77 (193)
10. Vocabulary Composite	.72 (198)	.72 (199)	.74 (198)
11. Comprehension Composite	.71 (194)	.71 (194)	.72 (193)

12. Sentence Comprehension	.70 (197)	.69 (197)	.70 (196)
13. Passage Comprehension	.64 (194)	.66 (194)	.67 (193)

Note. N = 200. Fifth-grade DIBELS Composite Score is comprised of Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. All correlations significant at $p < .001$

Table 50

BGS Sixth Grade DIBELS Composite Score Correlations with GRADE

Subtests and Composites	DCS by Time of Year		
	Beg.	Mid.	End
Raw Scores			
1. Total Test	.71 (103)	.68 (102)	.73 (105)
2. Vocabulary Composite	.71 (104)	.67 (103)	.71 (106)
3. Comprehension Composite	.64 (103)	.63 (102)	.68 (105)
4. Sentence Comprehension	.64 (104)	.60 (103)	.64 (106)
5. Passage Comprehension	.57 (103)	.57 (102)	.63 (105)
Standard Scores			
6. Total Test	.68 (103)	.67 (102)	.70 (105)
7. Vocabulary Composite	.69 (104)	.66 (103)	.69 (106)
8. Comprehension Composite	.59 (103)	.60 (102)	.63 (105)
Stanine Scores			
9. Total Test	.71 (103)	.69 (102)	.73 (105)
10. Vocabulary Composite	.72 (104)	.68 (103)	.72 (106)

11. Comprehension Composite	.61 (103)	.63 (102)	.65 (105)
12. Sentence Comprehension	.61 (104)	.59 (103)	.59 (106)
13. Passage Comprehension	.51 (103)	.52 (102)	.58 (105)

Note. N = 106. Sixth-grade DIBELS Composite Score is comprised of Retell, DORF Words Correct, a scaled score based on DORF Accuracy, and Daze Adjusted Score at all times of year. Students who did not have scores in all contributing measures for DIBELS were not given a composite score. Students who did not have all subtest scores for GRADE were not given a composite score, and students who did not have all composite scores for GRADE were not given a total test score. All correlations significant at $p < .001$

Determining the DIBELS Next Benchmark Goals and Cut Points for Risk Adequate

reading skills. The DIBELS Next benchmark goals provide targeted levels of skill that students need to achieve by specific times to be considered to be making adequate progress. In developing benchmark goals, our general focus is on adequate reading skills, and is not specific to a particular state assessment, published reading test, or national assessment. A student with adequate reading skills should read adequately regardless of the specific assessment that is used.

In the 2007 National Assessment of Educational Progress, 34% of students scored below the level of reading skills judged to be basic, and 68% of students scored below the level judged to be proficient. According to the NAEP, “basic denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at a given grade.” Thus, students who score at the 40th percentile or above on a high-quality, nationally norm-referenced test are likely to be rated Basic or above on the NAEP and can be considered to have adequate reading skills. In our benchmark goal study, we used the 40th percentile or above on the GRADE as *one approximation* of adequate reading skills. Our intent is to develop generalizable benchmark goals and cut points that will be relevant and appropriate for a wide variety of reading outcomes,

across a wide variety of states and regions, and for diverse groups of students. No single study can provide all the information necessary to evaluate generalizability. As with DIBELS 6th Edition, multiple studies will evaluate the reliability, validity, and utility of DIBELS Next. We are ultimately most interested in the convergence of evidence from many research studies that utilize many different sites, samples of students, and reading outcome measures.

GRADE as initial external criterion. We used the Group Reading Assessment and Diagnostic Evaluation (GRADE), a high-quality, nationally norm-referenced assessment, as an external criterion in our Benchmark Goal Study. We emphasized the GRADE Total Test Raw Score as the primary score to examine. In our analyses we found that the total score worked better as a criterion than the individual scores, and that the individual scores were related to other measures much the same as the total score was related to other measures. The lowest raw score on the GRADE that was at or above the 40th percentile compared to the GRADE normative sample was used as an approximation of the external criterion of adequate reading skills. The lowest raw score on the GRADE that was at or above the 20th percentile compared to the GRADE normative sample was used as an approximation for the external cut point for risk. Subsequent research will be essential to verify and replicate these findings with a range of other external criterion measures.

DIBELS Composite Score as Primary Internal Criterion.

We used the DIBELS Composite Score as a primary internal (i.e., within the DIBELS assessment system) criterion because it is the best indicator of the student's overall reading proficiency. This represents a change from the DIBELS 6th Edition procedures where DIBELS

ORF was used as the primary indicator of a student's reading proficiency. In our research with DIBELS Next, we find that, although the DIBELS ORF Words Correct score is very good in isolation, the DIBELS Composite Score is substantially better. For example, the end-of-year third-grade DORF Words Correct correlates .66 with the end-of-year GRADE Total Test Raw Score, which is a very strong validity coefficient. However, the end-of-year, third-grade DIBELS Composite Score correlates .75 with the end-of-year GRADE Total Test Raw Score, explaining 13% more variance than DORF alone. In general, we find that the DIBELS Composite Score provides a better overall measure of reading proficiency than the best single DIBELS Next measure at almost every grade and time of year. In addition to correlating more highly with external outcomes, the DIBELS Composite Score also provides a larger and more complete sample of reading behavior than any single measure in isolation. Thus, the DIBELS Composite Score serves as a very important internal criterion in developing and validating the DIBELS Next benchmark goals and cut points for risk.

Step-by-Step Procedures

The principle vision for DIBELS Next is a step-by-step vision. Student skills at or above benchmark at the beginning of the year put the odds in favor of the student achieving the middle-of-year benchmark goal. In turn, students with skills at or above benchmark in the middle of the year have the odds in favor of achieving the end-of-year benchmark goal. Finally, students with skills at or above benchmark at the end of the year have the odds in favor of adequate reading skills on a wide, general variety of external measures of reading proficiency.

Our fundamental logic for developing the benchmark goals and cut points for risk was to begin with the external outcome goal and work backward in that step-by-step system. We first obtained an external criterion measure (the GRADE Total Test Raw Score) at the end of the year

with a level of performance that would represent adequate reading skills. Next we specified the benchmark goal and cut point for risk on the end of year DIBELS Composite Score with respect to the end-of-year external criterion. Then, using the DIBELS Composite end-of-year goal as an internal criterion, we established the benchmark goals and cut points for risk on the middle-of-year DIBELS Composite Score. Finally, we established the benchmark goals and cut points for risk on the beginning-of-year DIBELS Composite Score using the middle-of-year DIBELS Composite Score as an internal criterion.

Once the benchmark goals and cut points for risk were established for the DIBELS Composite Score, they were used to establish the specific goals and cut points for risk for each individual DIBELS Next measure. The same step-by-step procedures were used for the individual measures.

Primary design specifications for benchmark goals and cut points for risk. The primary specification for the DIBELS Next benchmark goals was to establish a level of skill where students scoring at or above benchmark have favorable odds (80%-90%) of achieving subsequent reading outcomes. In other words, students scoring at or above the benchmark goal are in a zone where we are reasonably confident they will make adequate progress. The primary specification for a DIBELS cut point for risk is a level of skill where students scoring below that level have low odds (10%-20%) of achieving subsequent reading outcomes. In other words, students scoring below the cut point for risk are in a zone where we are reasonably confident the student will not make adequate progress unless provided with additional, intensive support.

In between the benchmark goal and the cut point for risk is a level of skill where the odds are about even (40%-60%) of achieving subsequent reading outcomes. We are not confident that students with skills in this range will make adequate progress; we are also not confident that they

will not. In other words, between the benchmark goal and the cut point for risk is a zone of uncertainty where we cannot make a good prediction of outcomes. By providing additional, strategic support to students with skills in this range along with progress monitoring, we can increase the likelihood that the student will make adequate progress.

Secondary design specifications for benchmark goals and cut points for risk. A

secondary consideration in establishing benchmark goals and cut points for risk was based on an examination of marginal percents. We tried to keep the marginal percent of students in each score level consistent from predictor to criterion. For example, 73% of students in our third-grade sample scored at or above the 40th percentile on the GRADE external criterion measure, indicating a fairly high performing sample. We set the third-grade end-of-year benchmark goal so that 73% of the sample also scored at or above benchmark on the DIBELS Next Composite Score. Thus, the sample appears equally high performing on both the DIBELS predictor and the GRADE criterion.

Another important secondary consideration in establishing benchmark goals and cut points for risk was based on the logistic regression predicting the odds of scoring at or above benchmark on the criterion based on their score on the predictor. For all students in the “At or Above Benchmark” range, the odds of achieving subsequent goals may be 80% to 90%; however, for students at the high end of that range the odds are somewhat higher, and for students at the low end of that range the odds are somewhat lower. The logistic regression analysis was used to estimate the odds of achieving subsequent early literacy goals for students who obtain the exact benchmark goal or the exact cut point for risk score. We tried to keep the predicted odds for students obtaining the exact benchmark goal at 60% or higher of achieving subsequent goals. We also tried to keep the predicted odds of achieving subsequent goals at 40%

or less for students obtaining the exact score corresponding to the cut point for risk. For example, on the third-grade end-of-year DIBELS Next assessment, the predicted odds of scoring at or above the 40th percentile on the GRADE were 67% for students scoring exactly the DIBELS Composite Score benchmark goal; the odds were 32% for students scoring exactly the cut point for risk.

Other design specifications for benchmark goals and cut points for risk. In addition to the primary and secondary considerations in establishing benchmark goals and cut points for risk, we also considered a number of issues including:

- The pattern of student performance in the scatterplot. We tried to establish goals where the majority of students scoring at or above benchmark on the predictor were also at or above benchmark in the criterion; where students who scored below benchmark on the predictor were equally split by the benchmark goal on the criterion; and where the majority of students who were below the cut point for risk were below the benchmark goal on the criterion.
- The receiver operator characteristic (ROC) curve analysis. A large area under curve (AUC) is desirable in ROC analysis and indicates a good trade-off of sensitivity and specificity. Benchmark goals in the upper left corner of the curve represent a balance of sensitivity and sensitivity.
- We also examined and considered other metrics for decision utility including sensitivity, specificity, negative predictive power, positive predictive power, percent accurate classification, and Kappa.
- Finally, we considered the overall pattern of benchmark goals and cut points for risk across measures and grades, and the historical benchmark goals and cut points for risk

from DIBELS 6th Edition. In addition, we considered the theoretical relations between core components of early literacy in our model.

Overall evaluative judgment. We specified the benchmark goals and cut points for risk as an overall evaluative judgment of primary, secondary, and other design specifications. No single concern was used in isolation from other concerns. Frequently we had to balance disparate concerns to obtain a satisfactory compromise. For example, increasing the benchmark goal might result in a better match of marginal percents, but might compromise the predicted odds in the logistic regression analysis. Alternatively, a lower benchmark goal might work better for the beginning of year to middle of year analysis, but perform more poorly in the middle of year to end of year analysis. In other cases, the logistic regression analysis did not fit the data well, and consequently the role of the logistic regression analysis was discounted in establishing the benchmark goals and cut points for risk. The benchmark goals and cut points for risk represent our best balance of all the considerations identified here. The benchmark goals and cut points for risk are presented in Figure 2.

Figure 2. *DIBELS Next* Summary of Benchmark Goals and Cut Points for Risk

DIBELS Composite Score									
26	122	119	113	130	155	141	190	238	
13	85	89	97	100	111	109	145	180	
First Sound Fluency (FSF)									
10	30								
5	20								
Letter Naming Fluency (LNF)									
No benchmark set for LNF									
Phoneme Segmentation Fluency (PSF)									
20	40	40							
10	25	25							
Nonsense Word Fluency (NWF)									
17	28	27	43	58		54			
8	15	18	33	47		35			
Correct Letter Sounds	Whole Words Read	1	8	13		13			
		0	3	6		6			
		DIBELS Oral Reading Fluency (DORF)							
Words Correct	Accuracy	23	47	52	72	87			
		16	32	37	55	65			
		78% 68%	90% 82%	90% 81%	96% 91%	97% 93%			
Retell Quality of Response	Retell	15	16	21	27				
		0	8	13	18				
				2	2	1			
Daze									

This page is adapted from a chart developed by Cache County School District.

Figure 2. DIBELS Next Summary of Benchmark Goals and Cut Points for Risk.

Each benchmark goal and cut point for risk is supported by one or more detailed analyses. These analyses include scatterplots, contingency tables, logistic regression analyses, receiver operator characteristic (ROC) curves, and summaries of other decision-utility metrics. These analyses are summarized in analysis detail pages (see Appendix A), which report how a predictor (or screening decision) variable is related to a criterion (or outcome) variable. These analysis detail pages include the following elements: heading, scatterplot, contingency table, logistic regression, ROC curve analysis, and a table of other decision utility indices. The results of these analyses for all measures by grade level and time of year can be found in Appendix A, beginning on page 207.

Heading

The heading at the top of each analysis detail page provides information about which two variables are examined. The predictor is provided first, along with the corresponding variable name, benchmark goal and cut point for risk, and description of the variable. Variable names are used in the benchmark goal analysis detail pages to summarize a large amount of information on one page per analysis. The criterion or outcome is provided next, also with the corresponding variable name, benchmark goal and cut point for risk, and a text description of the variable.

In Figure 3, the predictor or basis for a screening decision is the DIBELS Composite Score (identified by *DCS* in the variable name) from the third-grade (identified by the number 3 in the variable name), end-of-year (identified by the letter *e* in the variable name) assessment with a benchmark goal of 330 and a cut point for risk of 280. The criterion is the GRADE Total Test Raw Score (*gtotr*) from the third-grade (3), end-of-year (*e*) assessment with a benchmark goal of

83 and a cut point for risk of 71. The benchmark goal for the GRADE Total Test Raw Score (83) was specified to be the lowest raw score that was at or above the 40th percentile according to the GRADE norms. The corresponding cut point for risk (71) was the lowest raw score that was at or above the 20th percentile.

	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year

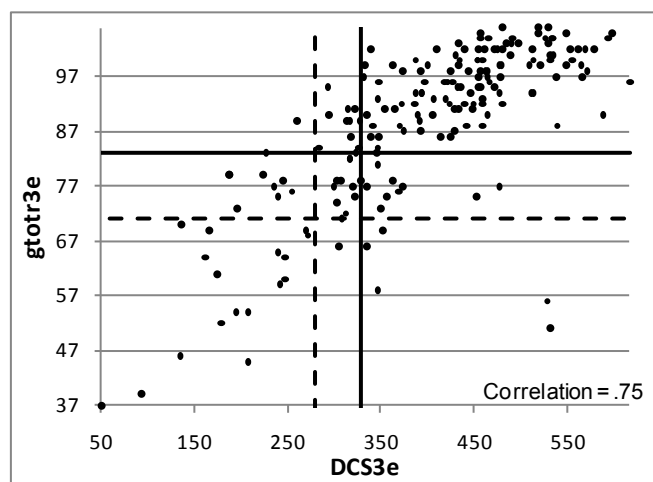
Figure 3. Benchmark goal analysis detail header.

Scatterplot

The scatterplot provides a visual representation of the relation between student scores on the predictor (along the horizontal axis) and the criterion (along the vertical axis). Each dot represents an individual student's scores on the predictor and criterion. The vertical solid line (on the right) represents the benchmark goal for the predictor where students performing at or above the benchmark goal are likely to need core support to achieve the benchmark goal on the criterion. The horizontal solid line (on the top) represents the benchmark goal on the criterion. The vertical dashed line (to the left) represents the cut point for risk on the predictor where students scoring below the cut point for risk are likely to need intensive support to achieve the benchmark goal on the criterion. The horizontal dashed line (on the bottom) represents the cut point for well below benchmark performance on the criterion. The correlation between the predictor and criterion is also reported in the lower right corner of the scatterplot.

In Figure 4, the scatterplot for DIBELS Composite Score for third grade end of year (DCS3e) and GRADE Total Test Raw Score for third grade end of year (gtotr3e) is portrayed. Two extreme outliers are evident in the lower right corner of the scatterplot who received very high scores on the DIBELS Composite, and very low scores on the GRADE. An educator would want

to follow up with those students to make sure an accurate estimate of their skills and progress is obtained. Even with those outliers, the two measures are correlated .75. Visually we see that most students who scored at or above the benchmark goal on DCS3e also scored at or above the benchmark goal on the GRADE. Most students who scored below the cut point for risk on the DCS3e also scored below the benchmark goal on the GRADE. Students who scored between the cut point for risk and the benchmark goal on the DCS3e were about evenly split, with about half above the benchmark goal on the GRADE and half below the goal.



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Figure 4. Scatterplot illustrating the relation between DIBELS Composite Score for third grade end of year (DCS3e) and GRADE Total Test Raw Score for third grade end of year (gtotr3e).

Contingency Table

The contingency table is presented directly below the scatterplot. The contingency table summarizes the number of students scoring in each zone of the scatterplot. The predictor variable forms the columns of the table and the criterion variable forms the rows. In this way, the contingency table corresponds directly to the zones of the scatterplot. The likely need for support

is used to label levels of performance on the predictor; score level is used to label levels of performance on the criterion. The marginal total number of students in each level of the predictor is provided in the first column margin; the marginal total number of students scoring in each level of the criterion is provided in the first row margin. Marginal percents of the total number of students are provided in the second margin. Below the table, the odds of students with that screening decision achieving the benchmark goal on the criterion are provided. The odds were obtained as a conditional percent of students achieving the criterion goal given that level of performance on the predictor.

For example, the contingency table for the third-grade, end-of-year DIBELS Composite Score as the predictor and the GRADE Total Test Raw Score also administered at end of year as the criterion is illustrated in Figure 5. The upper-left cell of the table, which corresponds to the upper-left zone of the scatterplot, indicates that there were two students who scored in the *Likely to Need Intensive Support* level of the predictor who also achieved the goal (*At or Above Benchmark*) on the GRADE external criterion measure. Overall, there were 27 students who scored in the *Likely to Need Intensive Support* range on the predictor, which was 14% of the total sample. That marginal percent is similar to the 12% of the sample that scored in the *Well Below Benchmark* range on the GRADE external criterion. Of the 27 students who were identified as *Likely to Need Intensive Support* on the predictor, only 2 (7%) achieved the goal. Thus, the odds were about 7% of achieving the benchmark goal for students with a screening decision of *Likely to Need Intensive Support*. For students who were identified as *Likely to Need Core Support* on the predictor, the odds of achieving the goal were 90%. For students who were identified as *Likely to Need Strategic Support*, the odds were 48%.

gtotr3e Outcome:	DCS3e Screening Decision:			Marginal total	Marginal percent
	Likely to need intensive support	Likely to need strategic support	Likely to need core support		
At or Above Benchmark	2	11	123	136	73%
Below Benchmark	8	11	9	28	15%
Well Below Benchmark	17	1	5	23	12%
Marginal Total	27	23	137	187	
Marginal Percent	14%	12%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	48%	90%		

Figure 5. The contingency table summarizes the number of students in each zone of the scatterplot, marginal totals, marginal percents, and the odds of students with a specific screening decision (e.g., *Likely to Need Intensive Support*) achieving the goal on the criterion.

Logistic Regression Analysis

The logistic regression analysis is provided directly to the right of the scatterplot. The small dots represent the moving percent (or likelihood) of students with that particular score on the predictor to achieve the benchmark goal on the outcome. To calculate those points, the students were rank-ordered from lowest to highest on the predictor score, and then overlapping intervals of students were created based on their scores on the predictor. The first interval started with the lowest score on the predictor, and went up to the score necessary to include at least 20 students. All students with a score in that range were included in the interval. For each subsequent interval, the lowest numeric score on the predictor was dropped, along with all students with that score, and the upper limit was raised until again at least 20 students were included in the interval. For example, the first interval might include all students with scores from 0 to 4 (if that was the minimum range necessary to include at least 20 students), and the second interval might include all students with scores from 1 to 7. Some intervals included more than 20 students because

multiple students obtained the maximum or minimum score for the interval. For each interval, the midpoint score on the predictor was identified and plotted on the horizontal axis. In the examples above, the midpoint of 0 to 4 is 2, and the midpoint of 1 to 7 is 4. The percent of students within the interval who achieved the benchmark goal on the outcome was calculated, and plotted on the vertical axis. The solid line connects the data points across intervals.

A logistic regression was performed predicting the odds of scoring at or above benchmark on the criterion based on the score on the predictor. The dashed line represents the logistic line of best fit to the data points. The large solid dot on the logistic regression line represents the predicted odds of achieving the goal for students who obtain the exact benchmark goal. The large open dot on the logistic regression line represents the predicted odds of achieving the goal for students who obtain the exact cut point for risk. Sometimes the logistic regression line provided a very good fit to the data and assisted in establishing benchmark goals and cut points for risk, but sometimes the model provided a poor fit to the data and was interpreted with caution.

For example, the third-grade, end-of-year DIBELS Composite Score and GRADE Total Test Raw Score logistic regression analysis is represented in Figure 6. The model fits fairly well and contributed to establishing benchmark goals and cut points for risk. Using the logistic regression model, the predicted odds of achieving the goal for a student exactly at benchmark on the predictor is 67%. The predicted odds of achieving the goal for a student exactly at the cut point for risk on the predictor is 32%.

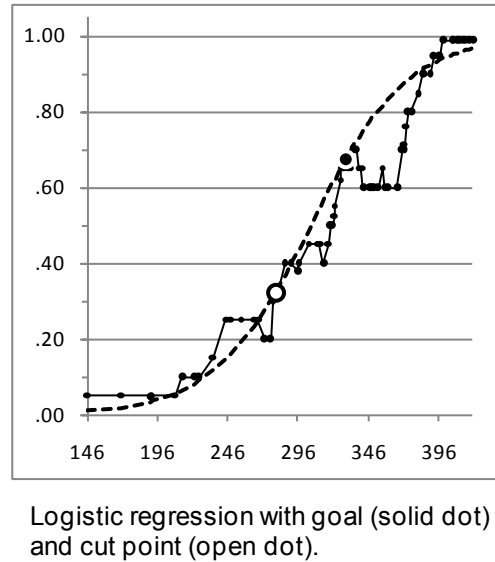


Figure 6. The logistic regression analysis summarizes the moving percent of students achieving the goal (solid line connecting small dots) and the logistic regression line fit to the moving percents (dashed line) with benchmark goal (large solid dot) and cut point for risk (large open dot).

Receiver Operator Characteristic Curve Analysis

The receiver operator characteristic (ROC) curve analysis is summarized directly to the right of the logistic regression analysis. The ROC curve is plotted by considering each possible score of the predictor as a potential decision point (either benchmark goal or cut point for risk). For each potential decision point, the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) are computed. True Positives refer to the number of students who were below the predictor score and who do not reach the goal (i.e., the screener indicated they would not reach the goal and was correct because they did not). False Positives refer to the number of students who were below the predictor score who did reach the goal (i.e., the screener or predictor indicated they would not reach the goal and the screener was in error because they did achieve the goal). Similarly, True Negatives are the number of students who

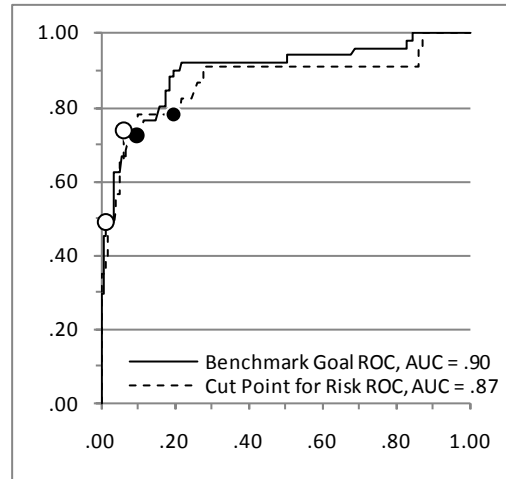
were above the predictor score who did achieve the goal, and False Negatives are the number of students who were above the predictor score but did not achieve the goal. The horizontal axis of the ROC curve is the False Positive Rate or $1 - \text{Specificity}$ calculated by $FP / (FP + TN)$. The vertical axis is the True Positive Rate or Sensitivity calculated by $TP / (TP + FN)$. In general there is a trade-off of sensitivity and specificity: as higher scores are considered for the decision rule, the sensitivity of the decision increases but the specificity declines. When the curve extends higher into the upper-left corner of the graph and the area under the curve (AUC) increases, there is a more favorable trade-off of sensitivity and specificity.

We should note here that we are troubled by the terminology of True Positives and False Positives. The premise of DIBELS is that there are no True Positives. In biomedical research, True Positive refers to subjects who were identified as likely to have the condition on the screening test (for example, they screened positive for tuberculosis) and who were then determined to actually have the condition (i.e., they were true positives). However, in an educational context, students who are at risk for later difficulty are provided with additional support for the explicit purpose of *changing* the outcome. Thus, in an educational context, students who are labeled as True Positives are actually students for whom an effective intervention was not provided. When calculating sensitivity and specificity in biomedical research, a False Positive represents an error in the screening decision. For example, a subject might be screened as positive for tuberculosis, but then determined not to have the condition (i.e., a false screening decision of positive). However, in an educational context, a student who is classified as a False Positive may not represent a screening error. The student may, in fact, have been accurately identified as at risk for a poor outcome, and been provided with an effective intervention that actively changed the outcome. As educators, it is our charge to ruin predictions

of difficulty. Similarly, it is not possible to distinguish between a False Negative (a student who is identified as needing core support, but who does not reach future benchmark goals) and an ineffective core curriculum or instruction.

In the context of DIBELS Next, we are interested in two levels of performance on the outcome or criterion measure: *At or Above Benchmark* where we are reasonably confident the student is making adequate progress, and *Well Below Benchmark* where we are reasonably confident the student is not making adequate progress. Thus, we compute two ROC curves for each analysis: one representing the benchmark goal on the criterion measure (solid line) and a second representing the cut point for risk on the criterion (dashed line). We also are interested in the sensitivity and specificity of two particular predictor scores: the benchmark goal on the predictor (large solid dot), and the cut point for risk on the predictor (large open dot).

For example, the ROC curves for the third-grade, end-of-year DIBELS Composite Score as predictor with respect to the GRADE Total Test Raw Score at end-of-year third grade is illustrated in Figure 7. The curves are both similar, and toward the upper-left corner of the graph with AUCs of .90 and .87, indicating an excellent trade-off of sensitivity and specificity. The benchmark goal is in the upper-left corner of the curve, indicating a good balance of sensitivity and specificity for that decision point. The cut point for risk on the benchmark goal ROC is lower on the curve, indicating a less than optimum balance of sensitivity and specificity for that decision point.



Receiver Operator Characteristic (ROC) curves.

Figure 7. The receiver operator characteristic (ROC) curve analysis summarizes the trade off of sensitivity (vertical axis) and specificity (1 – specificity on the horizontal axis).

Other Decision-Utility Indices

The final section of the benchmark goal analysis detail is a table summarizing other decision-utility indices that were considered. The table of other indices is located in the lower right corner of the analysis detail page. These decision-utility indices were developed primarily for 2-by-2 decisions: 2 levels on the predictor (at risk/positive or not at risk/negative) and 2 levels on the criterion (above goal/negative or below goal/positive). For DIBELS Next, we are interested in 3 levels on the predictor and 3 levels on the criterion. We handled this distinction by constructing and evaluating four 2-by-2 decision tables: (a) above or below benchmark on the criterion and above or below benchmark on the predictor, (b) above or below benchmark on the criterion and above or below the cut point for risk on the predictor, (c) above or below cut point for risk on the criterion and above or below benchmark on the predictor, (b) above or below cut point for risk on the criterion and above or below the cut point for risk on the predictor.

For example, the other decision-utility indices for the third-grade, end-of-year DIBELS Composite Score as predictor with respect to the GRADE Total Test Raw Score at end-of-year third grade are summarized in Figure 8. The greatest sensitivity is obtained when the predictor decision is whether the student is above or below the benchmark goal and the criterion outcome is whether the student is above or below the cut point for risk (.78). The greatest specificity is obtained when the predictor decision is whether the student is above or below the cut point for risk and the criterion outcome is whether the student is above or below the benchmark goal (.99). The percent accurate classification ranged from .80 to .91, and Kappa coefficients ranged from .39 to .63.

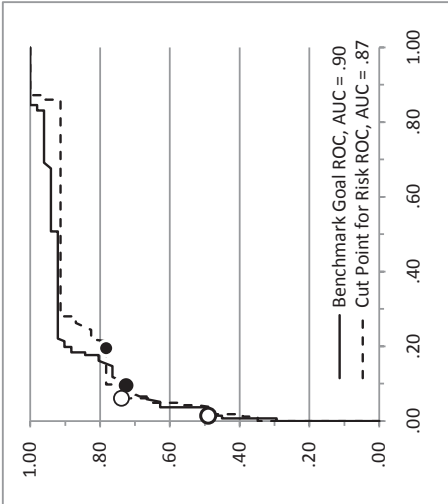
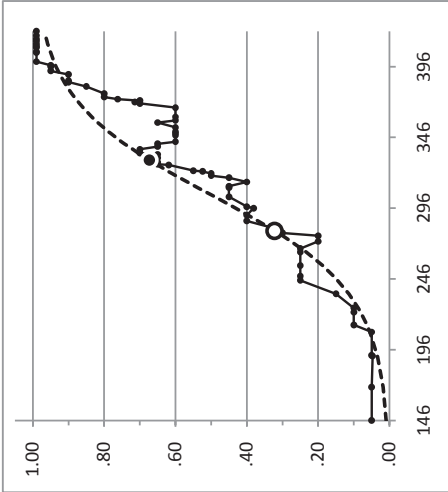
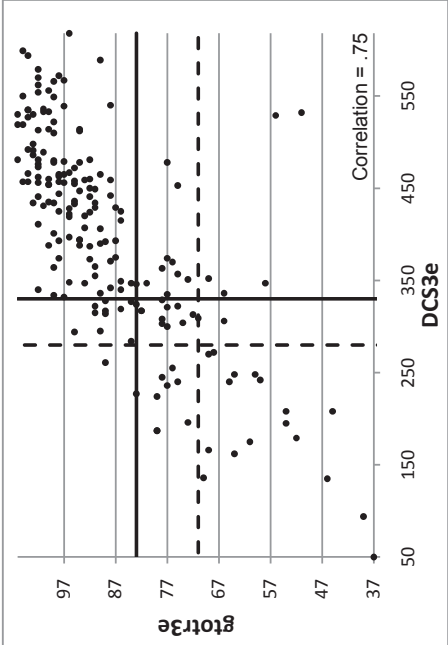
	<u>At or Above Benchmark outcome</u>		<u>Well Below Benchmark outcome</u>	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	134	132	154
False Negative	14	26	5	6
True Positive	37	25	18	17
False Positive	13	2	32	10
Sensitivity	.73	.49	.78	.74
Specificity	.90	.99	.80	.94
Negative Predictive Power	.90	.84	.96	.96
Positive Predictive Power	.74	.93	.36	.63
Accurate Classification	.86	.85	.80	.91
Kappa	.63	.56	.39	.63

Figure 8. Other decision-utility indices.

Our primary design specification for setting benchmark goals is that a Low Risk Decision on initial DIBELS Next assessment should result in the favorable likelihood, or odds, (85% +/- 5%) of achieving subsequent reading health outcomes. In other words, this decision identifies a zone where we are reasonably confident the student has adequate skills. In addition, a Some Risk Decision on initial DIBELS Next assessment should result in 50 – 50 odds (50% +/- 10%) of achieving subsequent reading health outcomes. In other words, this identifies a zone of

uncertainty where we don't know if the student is on track or not. Finally, a High Risk Decision on initial DIBELS Next assessment should result in low odds (15% +/- 5%) of achieving subsequent reading health outcomes – unless intensive intervention is implemented. In other words, this decision identifies a zone where we are reasonably confident the student does not have adequate skills.

Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year
	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3 End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).

Logistic regression with goal (solid dot) and cut point (open dot).

Receiver Operator Characteristic (ROC) curves.

DCS3e Screening Decision:

gtotr3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support
At or Above Benchmark	2	11	123
Below Benchmark	8	11	9
Well Below Benchmark	17	1	5
Marginal Total	27	23	137
Marginal Percent	14%	12%	73%

Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)

7% 48% 90%

At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	134	132	154
False Negative	14	26	5	6
True Positive	37	25	18	17
False Positive	13	2	32	10
Sensitivity	.73	.49	.78	.74
Specificity	.90	.99	.80	.94
Negative Predictive Power	.90	.84	.96	.96
Positive Predictive Power	.74	.93	.36	.63
Accurate Classification	.86	.85	.80	.91
Kappa	.63	.56	.39	.63

Figure 9. Benchmark Goal Detail for Third Grade End of Year DCS with end of year GRADE Total Test.

Assessor and Teacher Satisfaction with DIBELS Next Measures

DIBELS BGS Usability Questionnaire data were examined in three ways. First, descriptive statistics (e.g., means, standard deviations) were calculated for each question. These data are displayed in Tables 51 and 52 for the participants who completed the questionnaire. Some participants did not respond to some items, resulting in a reduced sample size for some of the items as shown in Table 51. In addition, descriptive statistics are reported along with the results of the second analysis.

In the second analysis, the percent of participants responding to each rating on each item was examined. Results of this analysis are reported by question in a series of figures. Finally, our third analysis involved examining all of the narrative feedback obtained via the questionnaire. Themes found in the responses are reported by item (question #).

Item 4: *The administration and scoring rules for the next version of DIBELS are easy to follow.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.37 indicating agreement with this statement. The majority of respondents (54%) indicated they agreed, and 39% indicated strong agreement with this statement.

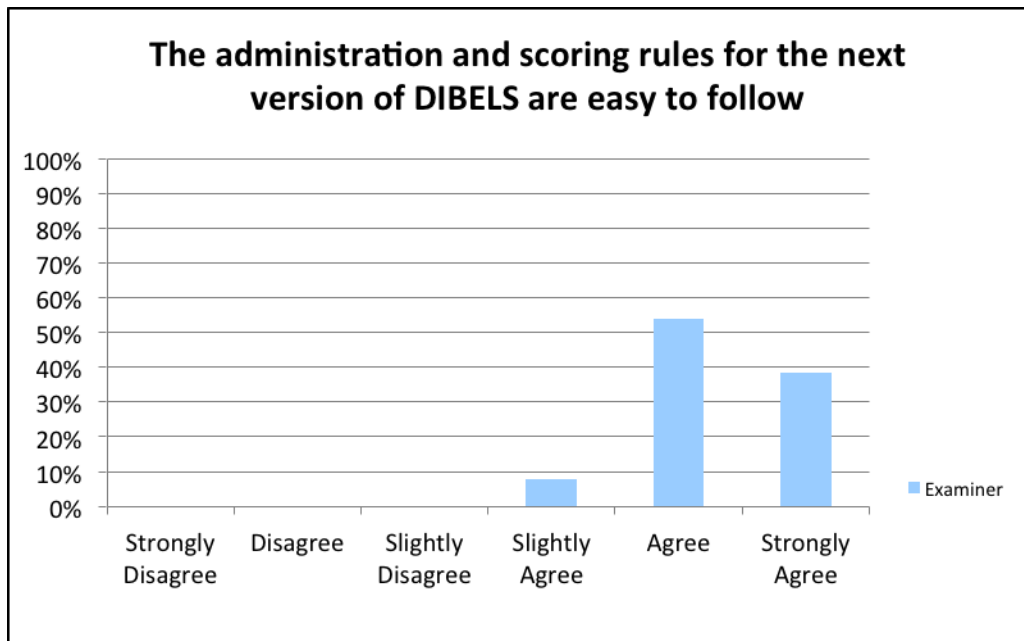


Figure 10. Percent of participant agreement with item 4 of the DIBELS BGS Usability Questionnaire.

Item 6: *The next version of DIBELS materials are organized for efficient administration.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.04 indicating agreement with this statement. The majority of respondents (65%) agreed with this statement. About 19% strongly agreed with this statement.

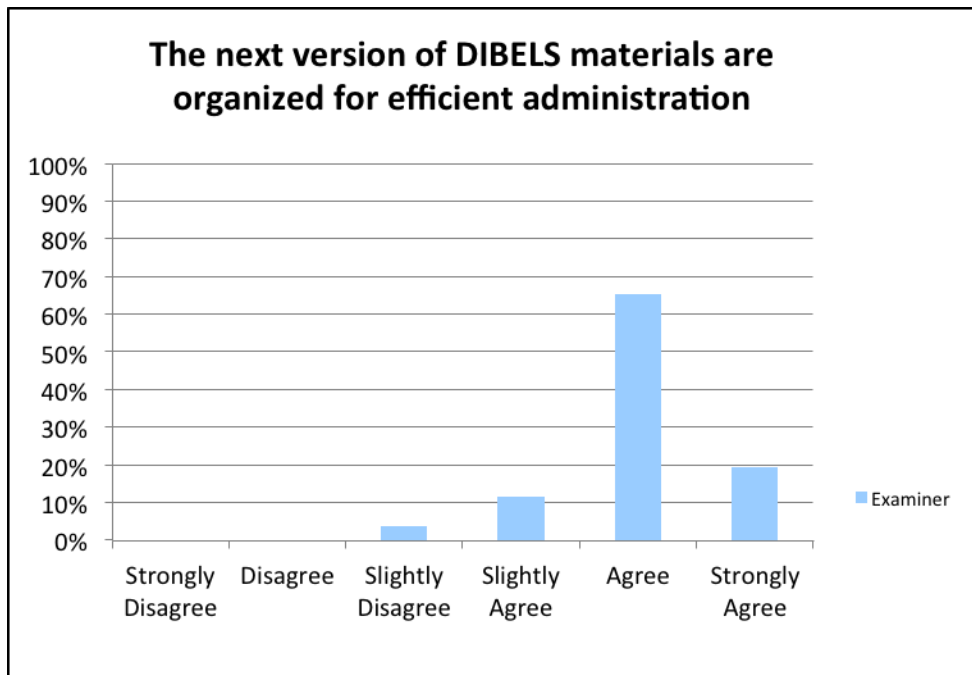


Figure 11. Percent of participant agreement with item 6 of the DIBELS BGS Usability Questionnaire.

Item 8: *The wording of directions in the next version of DIBELS measures ensures students understand the tasks.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.14, indicating general agreement with this statement. Once again, the largest percentage of responses (69%) agreed with this statement, but with 19% strongly agreeing with the statement.

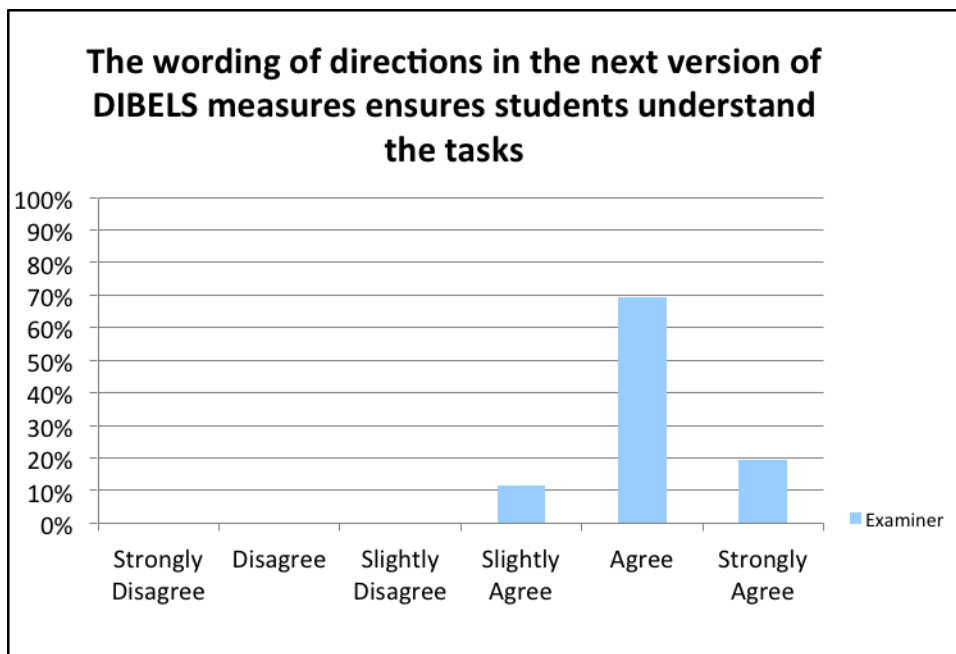


Figure 12. Percent of participant agreement with item 8 of the DIBELS BGS Usability Questionnaire.

Item 10: *The number, type, and sequence of practice items on the next version of DIBELS are sufficient to ensure that students understand the tasks.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.22 indicating general agreement with this statement. The largest percentage of responses were for the "Agree" (73%) and "Strongly Agree" (23%) ratings.

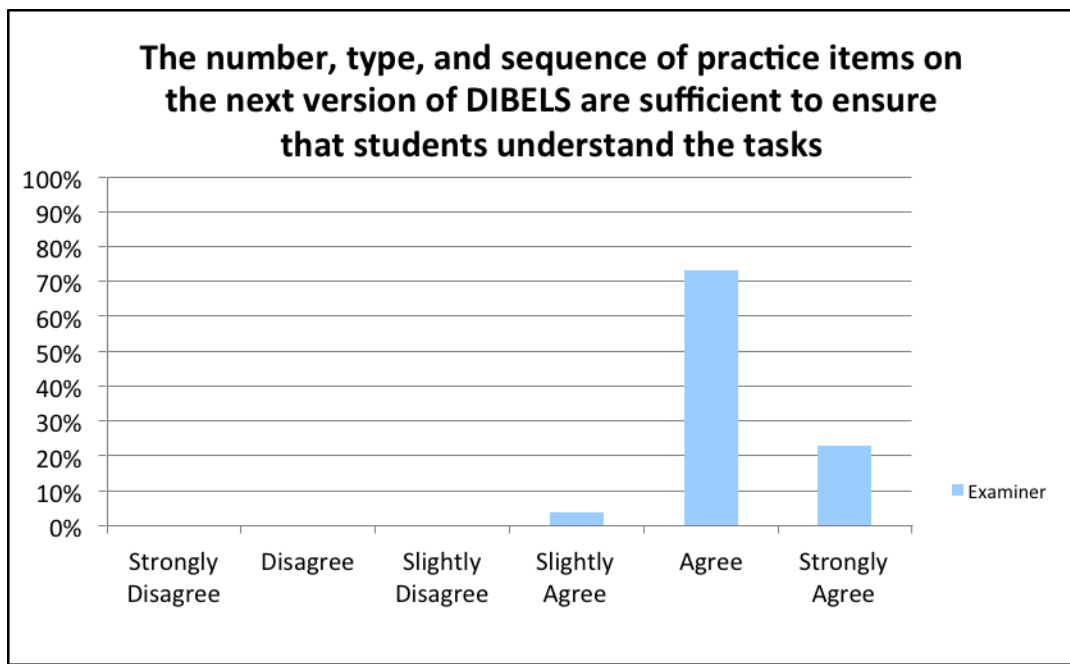


Figure 13. Percent of participant agreement with item 10 of the DIBELS BGS Usability Questionnaire.

Item 12: *The new administration procedures and scoring rules for the next version of DIBELS do not cause any unexpected difficulties during testing.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.04 indicating general agreement with this statement. The largest percentages of responses were in the "Agree" range (69%) with the remainder of responses being divided about equally between "Somewhat Agree" and "Strongly Agree."

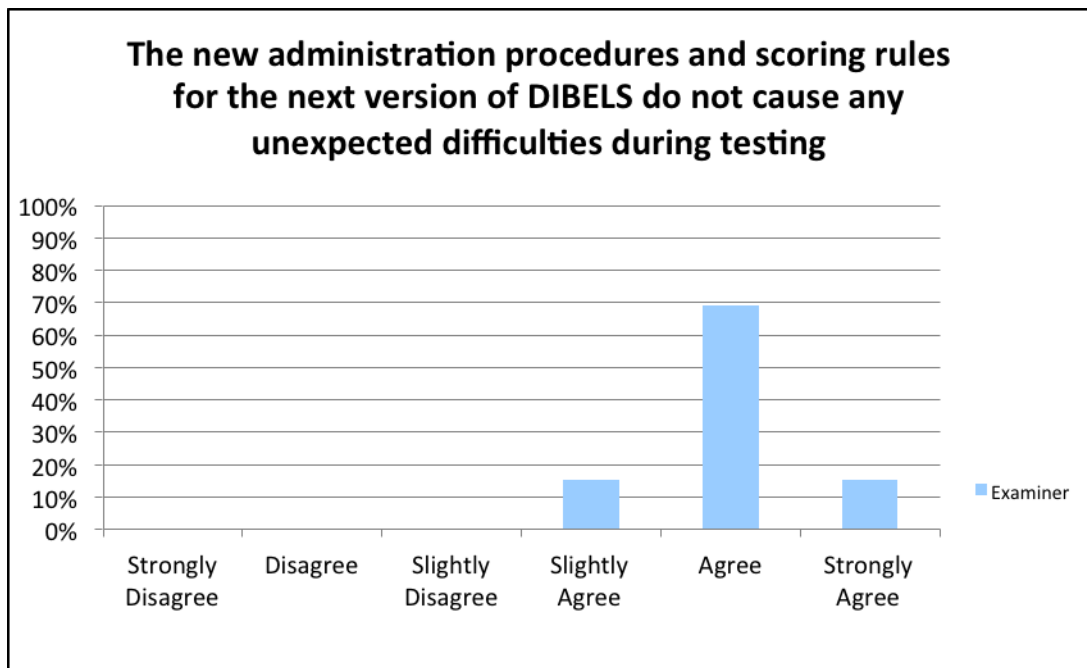


Figure 14. Percent of participant agreement with item 12 of the DIBELS BGS Usability Questionnaire.

Item 14: *The next version of DIBELS score sheets make it easy to track items, and record a final score.*

Twenty-five participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.92 indicating general agreement with this statement. The largest percentage of responses were for the "Agree" (68%) and "Slightly Agree" (16%) ratings.

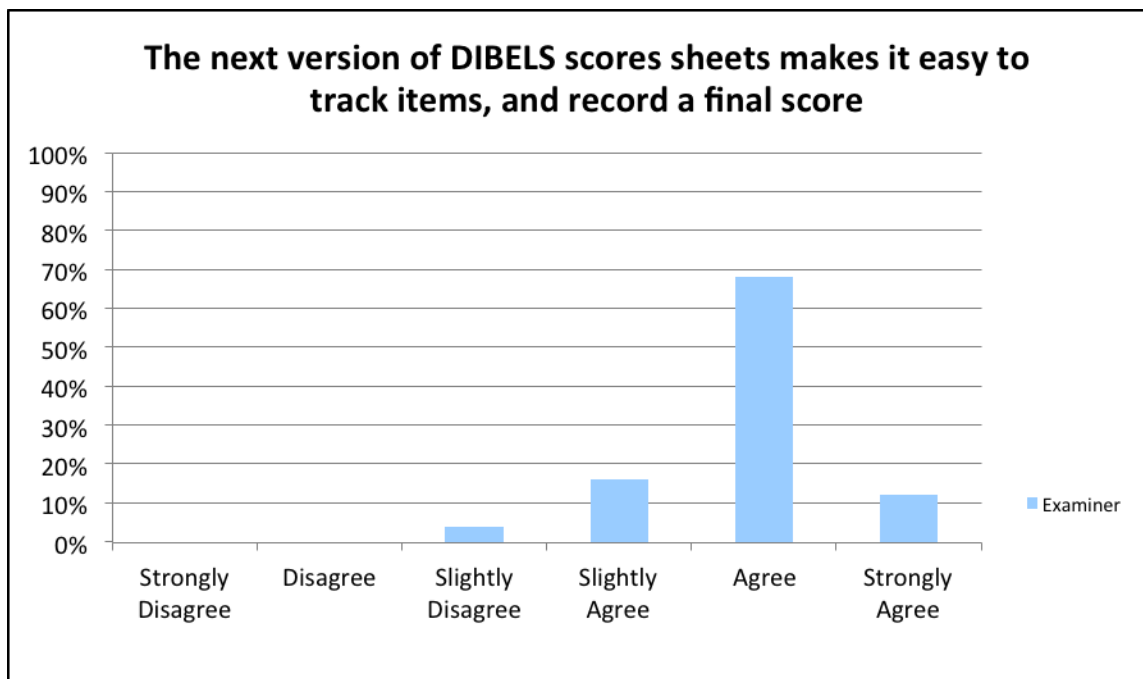


Figure 15. Percent of participant agreement with item 14 of the DIBELS BGS Usability Questionnaire.

Item 16: *The scores obtained from the next version of DIBELS are accurate indicators of students' development of critical early literacy skills.*

Sixteen participants from the "both teacher and assessor" group responded to this item. The mean rating was 4.63, indicating general agreement with this statement, but a slightly lower rating than previous items. Still, the majority of responses were "agree (44%), but about 38% were "Slightly Agree."

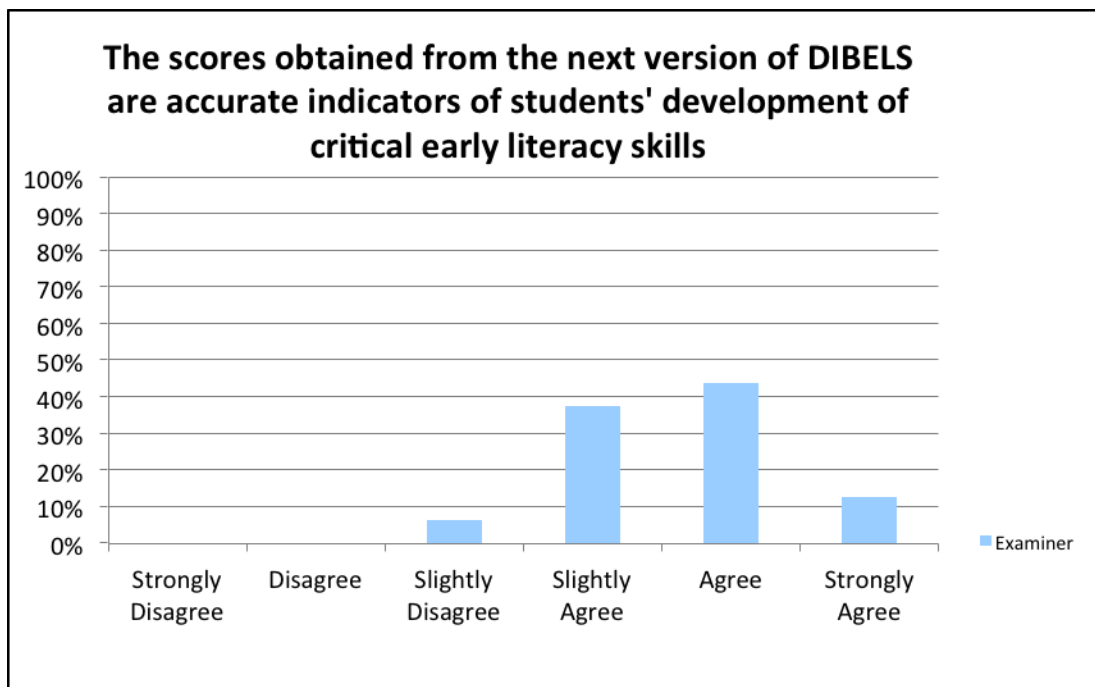


Figure 16. Percent of participant agreement with item 16 of the DIBELS BGS Usability Questionnaire.

Item 18: *The opportunity to indicate Response Patterns for the next version of DIBELS measures provides valuable information in addition to the scores.*

Twenty-three participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.56 indicating general agreement with this statement, but again slightly lower agreement than previous questions. The largest percentage of responses were for the "Agree" (52%) and "Slightly Agree" (35%) ratings.

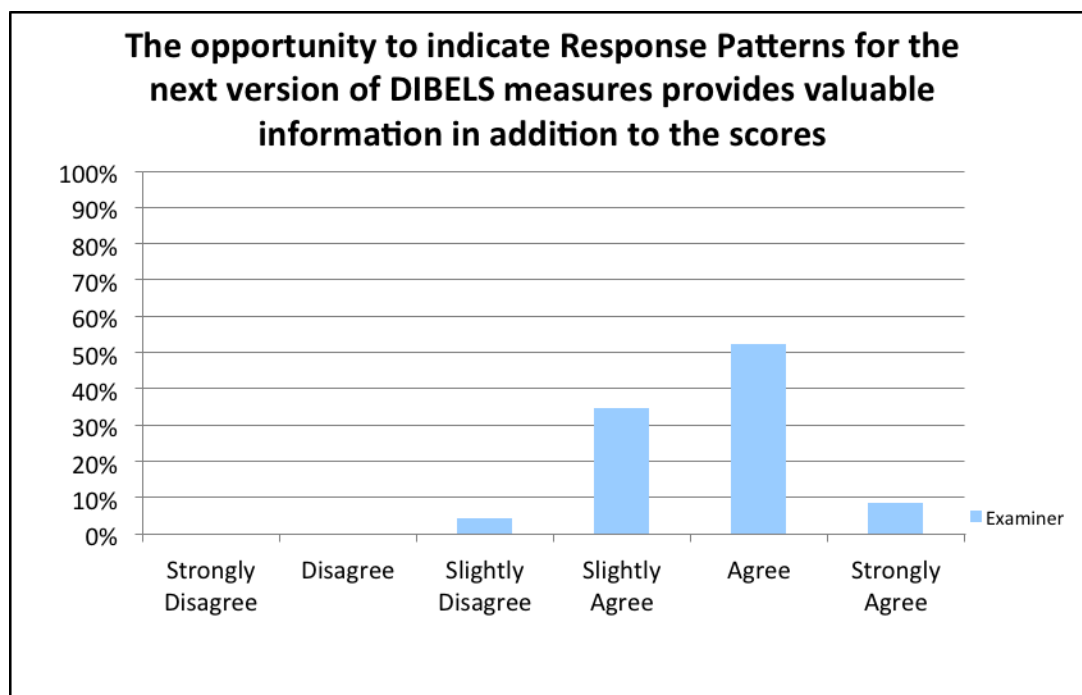


Figure 17. Percent of participant agreement with item 18 of the DIBELS BGS Usability Questionnaire.

Item 20: *The Quality of Response rating for the retell component of ORF provides valuable information in addition to the ORF/retell score.*

Twenty-four participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.48 indicating general agreement with this statement, but slightly lower agreement than previous questions and with greater variability in the responses. The largest percentage of responses were equally divided between the "Agree" and "Slightly Agree" ratings with 37.5% for each of these ratings. One person (4.2%) strongly disagreed with this statement, while 16.6% strongly agreed with this statement. In addition to the ratings responses, eleven respondents (46%) also provided narrative comments on this item. Two of the eleven mentioned the subjective nature of the Quality of Response rating and two others mentioned that teachers do not use this rating.

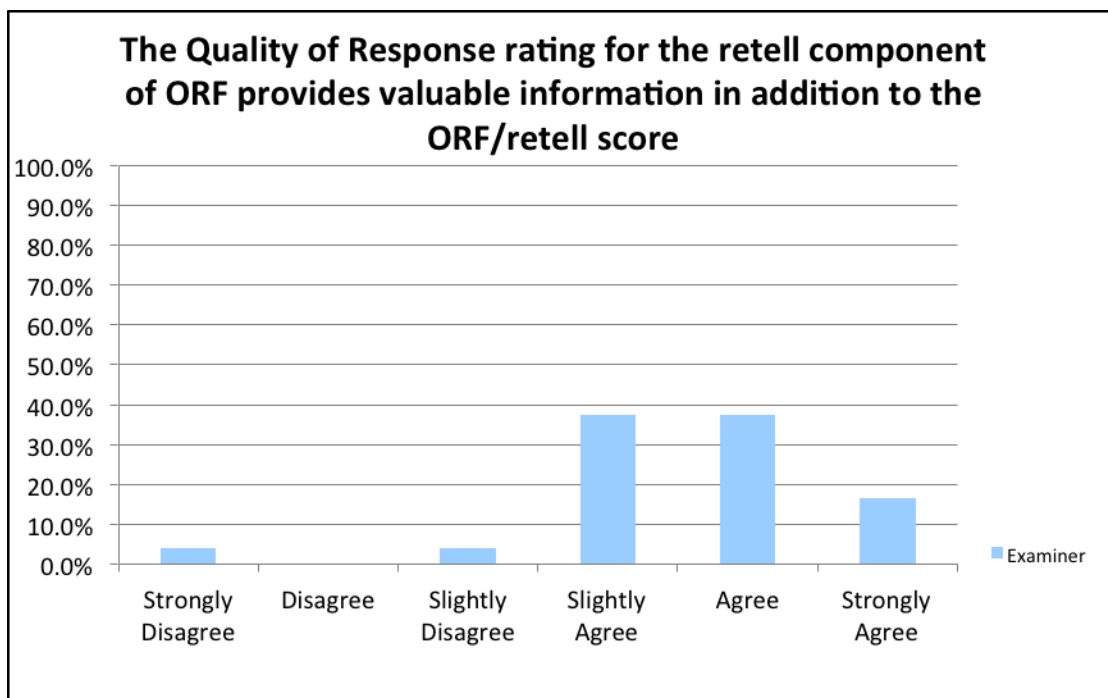


Figure 18. Percent of participant agreement with item 20 of the DIBELS BGS Usability Questionnaire.

Item 21: *The Quality of Response rating for WUF-R provides valuable information in addition to the WUF-R score.*

Twenty-three participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.53 indicating general agreement with this statement, but continuing the pattern of slightly lower agreement than the initial questions. The largest percentage of responses were for the "Agree" (52%) and "Slightly Agree" (35%) ratings. Five of the respondents (22%) also provided comments for this item. The only remark that occurred more than one time was related to the subjective nature of scoring Quality of Response (two respondents).

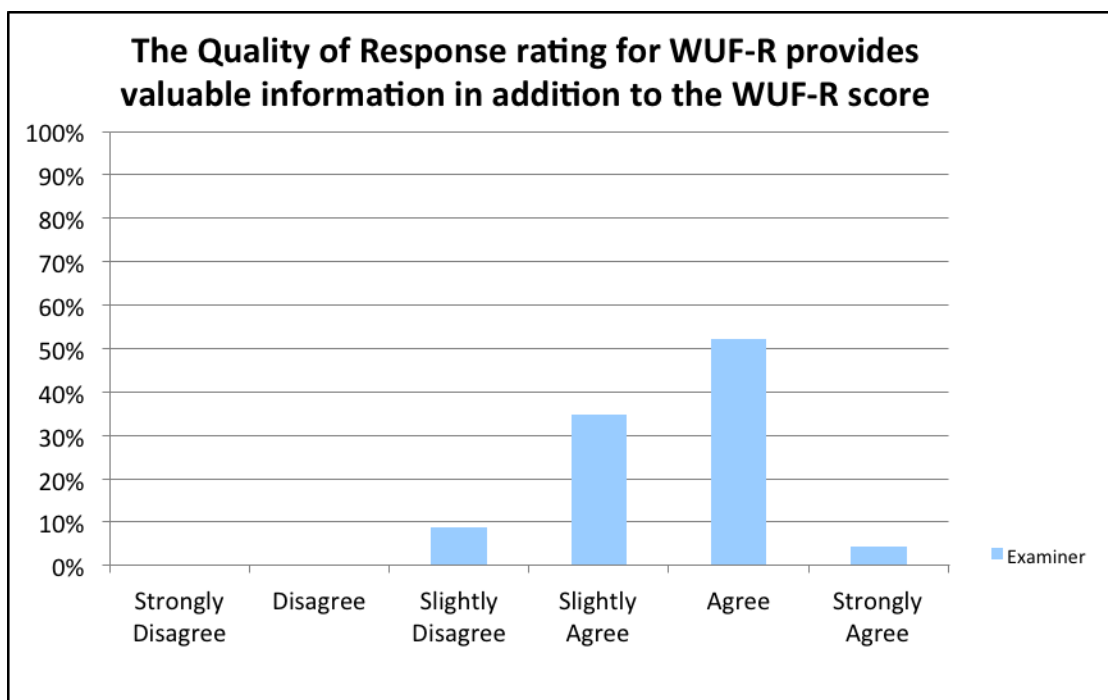


Figure 19. Percent of participant agreement with item 21 of the DIBELS BGS Usability Questionnaire.

Item 22: *Overall, data obtained from the next version of DIBELS are helpful for goal setting.*

Ninety-one participants from the "teachers only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.66 indicating general agreement with this statement. Like many of the other items, the largest percentage of responses were for the "Agree" (55%) and "Slightly Agree" (19%) ratings, but unlike other items, this time responses occurred in each rating category.

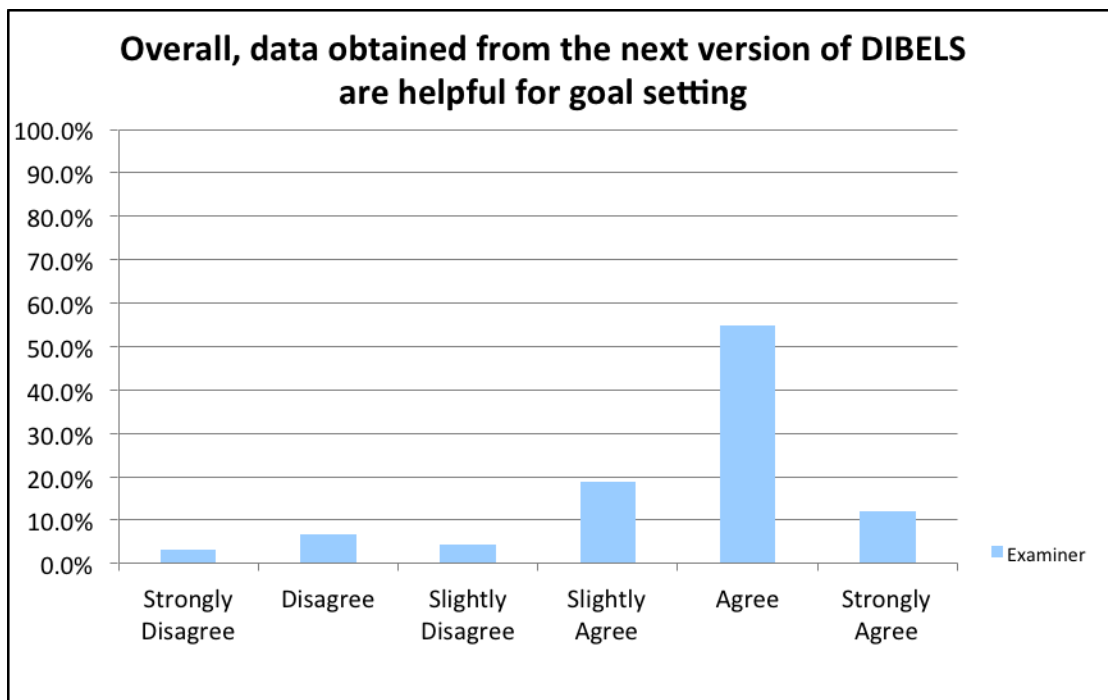


Figure 20. Percent of participant agreement with item 22 of the DIBELS BGS Usability Questionnaire.

Item 24: *Overall, data obtained from the next version of DIBELS are helpful for making instructional grouping decisions.*

Eighty-nine participants from the "teachers only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.81 indicating general agreement with this statement. The largest percentage of responses were for the "Agree" (61%) and "Slightly Agree" (15%) ratings, but once again responses occurred in each rating category.

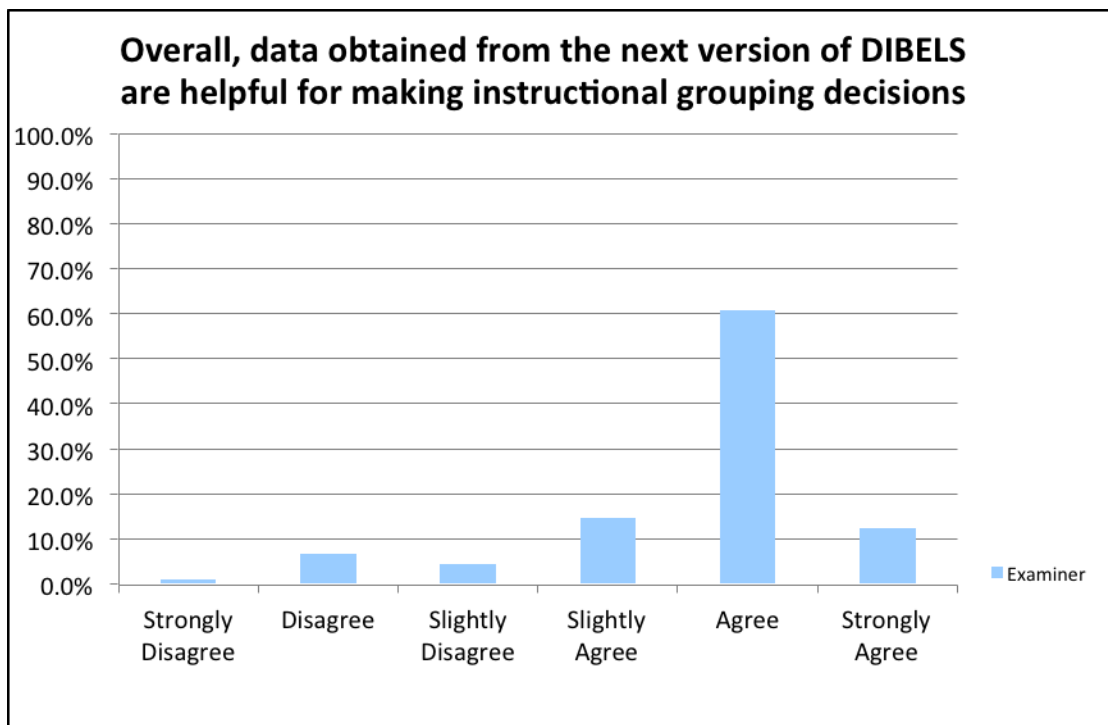


Figure 21. Percent of participant agreement with item 24 of the DIBELS BGS Usability Questionnaire.

Item 26: *Overall, data obtained from the next version of DIBELS are helpful for identifying instructional targets.*

Eighty-six participants from the "teachers only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.76 indicating general agreement with this statement. The largest percentage of responses were for the "Agree" (54%) and "Slightly Agree" (22%) ratings, and once again responses occurred in each rating category.

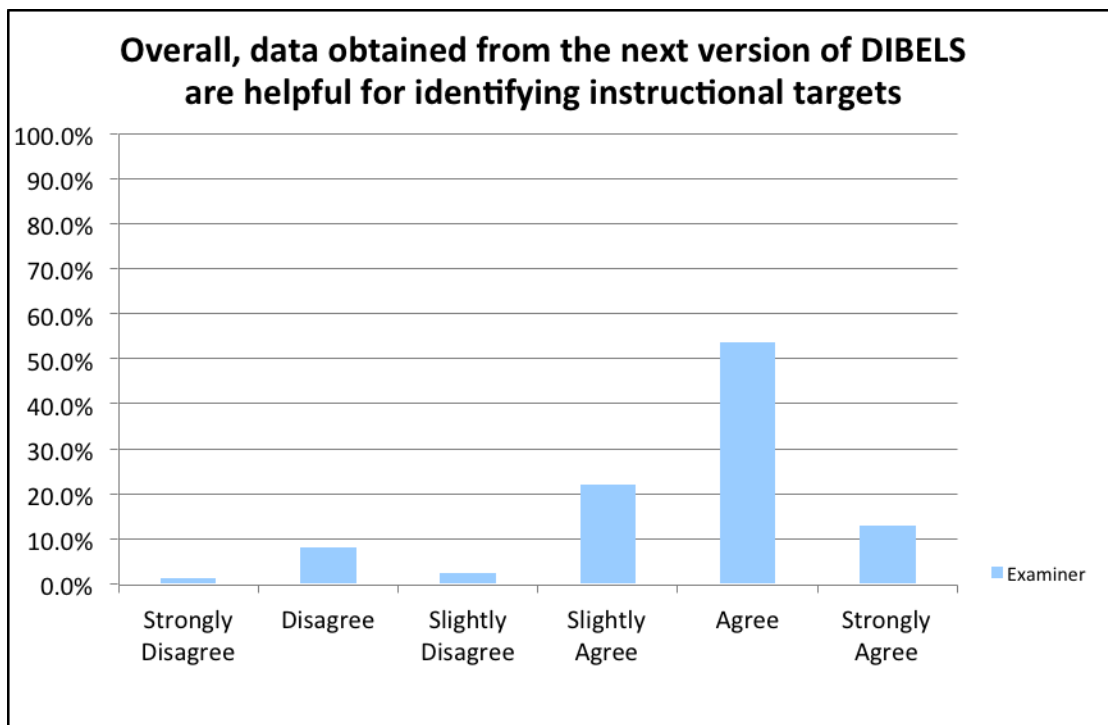


Figure 22. Percent of participant agreement with item 26 of the DIBELS BGS Usability Questionnaire.

Item 28: *The next version of DIBELS benchmark booklets are useful for efficient test administration and management of student scores.*

Twenty-six participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.06 indicating agreement with this statement. The largest percentage of responses were for the "Agree" (61.5%) and "Strongly Agree" (23%) ratings. Seven respondents (27%) also commented on this item. Two of these respondents mentioned that less page flipping/having scoring on the same page would be helpful. No other comments were made by more than one respondent for this item.

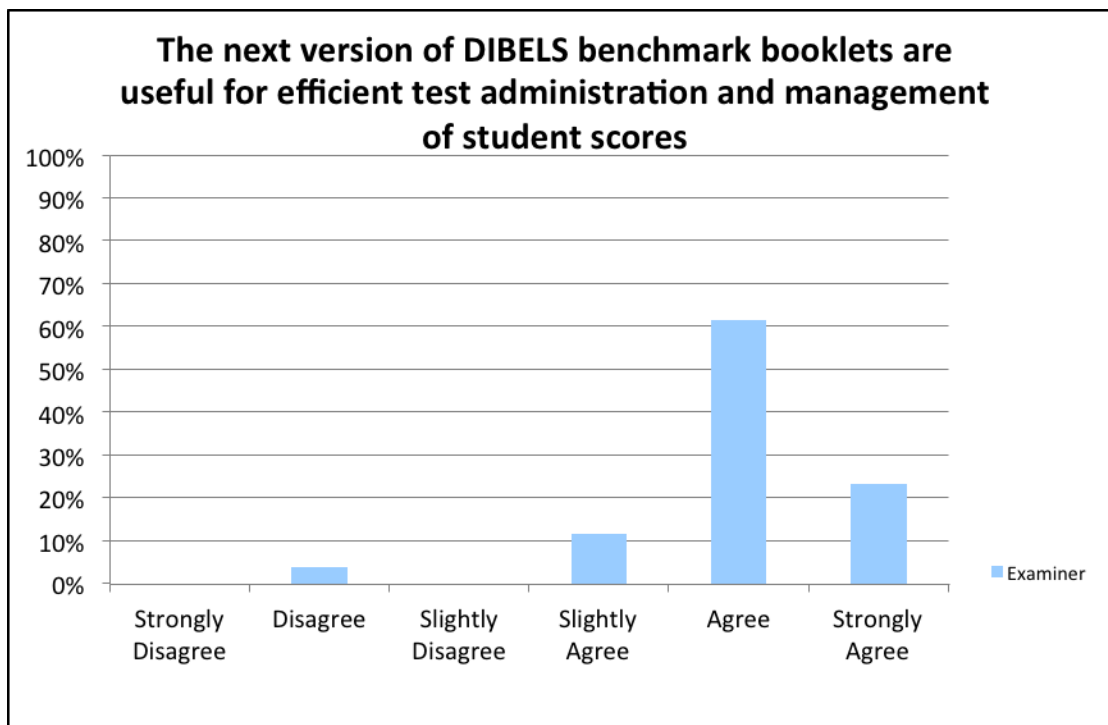


Figure 23. Percent of participant agreement with item 28 of the DIBELS BGS Usability Questionnaire.

Item 29: *The new formatting features for the next version of DIBELS materials do not cause any unexpected difficulties during testing.*

Twenty-five participants from the "assessor only" and "both teacher and assessor" groups responded to this item. The mean rating was 5.05 indicating agreement with this statement. The largest percentage of responses were for the "Agree" (64%) and "Strongly Agree" (20%) ratings.

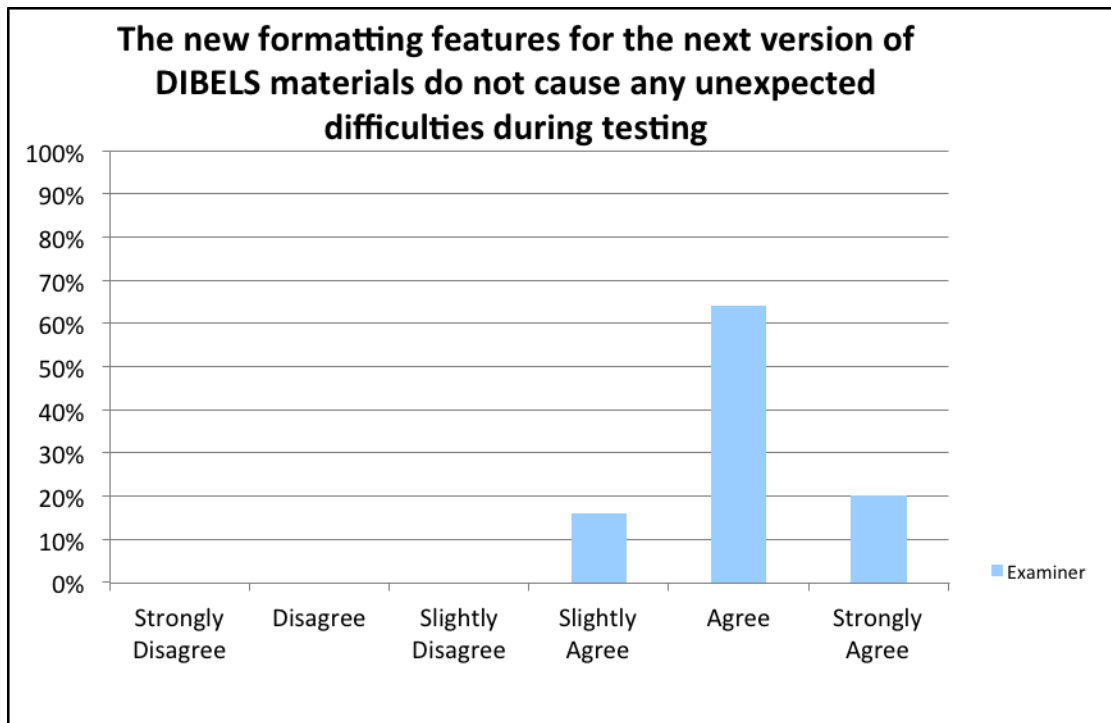


Figure 24. Percent of participant agreement with item 29 of the DIBELS BGS Usability Questionnaire.

Item 32: *I would suggest use of the next version of DIBELS to others.*

Eighty-four participants from the "teachers only" and "both teacher and assessor" groups responded to this item. The mean rating was 4.62 indicating general agreement with this statement. The largest percentage of responses were for the "Agree" (56%) and "Slightly Agree" (17%) ratings, and once again responses occurred in each rating category. Seventeen (20%) of the 84 respondents also commented on this item. Comments occurring more than once were: (a) use with other information (4 respondents); (b) not diagnostic/too narrow (3 respondents); and (c) time consuming (2 respondents).

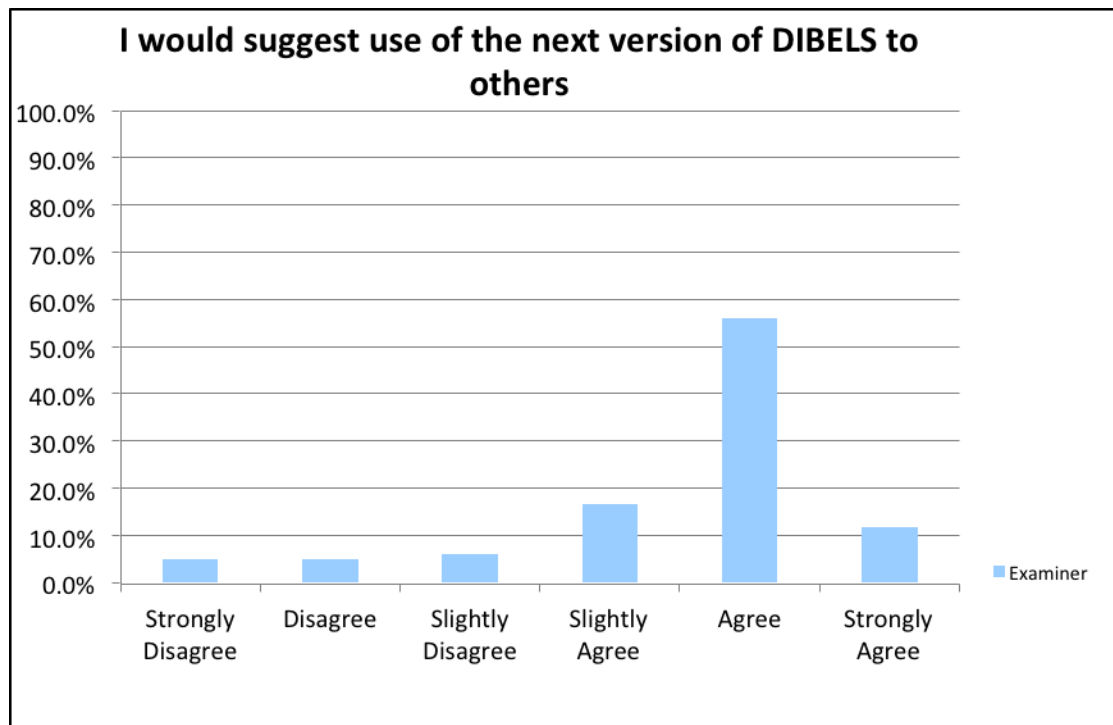


Figure 25. Percent of participant agreement with item 32 of the DIBELS BGS Usability Questionnaire.

Item 62: *Scores from the next version of DIBELS are consistent with other information I have about students in my class.*

Seventy participants from the "teachers only" group responded to this yes or no response item. Sixty respondents indicated "yes" (86%) and ten respondents (14%) responded "no."

Twenty-five of the respondents provided comments on this item. Comments centered around the theme of DIBELS scores being generally consistent with other information, but not always.

Third and finally, responses to specific open-ended items on the questionnaire were examined for themes. There were 17 open-ended items, many of which directly related to a specific rated item. For example, the open-ended item #5, "Please comment on the administration and scoring rules of individual measures," was directly related to item #4. All open-ended items directed respondents to comment on each individual measure. Test items and themes are reported in Table 53. Themes were identified if a comment was shared by at least two respondents.

In reviewing the data in Table 53, some items (#s 11, 30, 31, & 34) had no identified themes (i.e., no comment made more than once). Other items had common remarks made about each measure. For example, when asked to comment about the administration and scoring of each measure (item #5) themes found across measures suggest that respondents found the measures easy to administer and score. Other items where this were found were items #s 23, 25 and 27 with respect to the usefulness of individual measures for goal setting, making grouping decisions, and identifying instructional targets. The overarching theme across all three items was that the measures were useful/helpful or very useful/helpful for each of these decisions. Typically, twice as many (or more) respondents noted the utility of the measures rather than a lack of utility. Notably, across all three of these items, more explicit statements were provided

for NWF and DORF about how the measures were used, when compared to responses for the other measures. Some sample responses included:

- *I think it helps to see if the students are just saying the sounds or the whole words. It is helpful to push the students to get the higher skill (reading the whole word) and set goals to get to that skill.*
- *It is easy to group students using NWF because you can break it into different skills: 1) Not being able to say the sounds at all, 2)Saying just sounds and not any words, 3) Transitioning between sounds and whole words or part of words, 4) Saying the whole word.*
- *I am able to set goals for individual students when I see struggles in fluency, accuracy, and comprehension.*
- *I use this to group them in my class for their fluency levels.*
- *helps me focus my lesson planning on what the child really needs*

Another item for which common themes were noted across measures was item #33, "Please comment on your willingness to recommend the individual measures." Comments support a willingness to recommend all the measures, but DORF and NWF in particular had numerous comments indicating a willingness to recommend. Common themes were also found for each measure for item #63, "Please comment on the consistency of scores from individual measures with other information you have about students in your class." The predominant theme was that the measures were consistent with other information teachers had about their students, which is consistent with the responses to item #62.

Finally, two comments appeared several times across items. One of these issues was about page-flipping during the administration and scoring and the desire to have the scoring section be on the same page. The second comment that appeared several times had to do with the number of pages in DORF and passage length.

Table 51. *DIBELS BGS Usability Questionnaire Ratings*

Item	N	Mean Rating (SD)
4. The administration and scoring rules for the next version of DIBELS are easy to follow	26	5.37 (.57)
6. The next version of DIBELS materials are organized for efficient administration	26	5.04 (.74)
8. The wording of directions in the next version of DIBELS measures ensures students understand the tasks	26	5.14 (.51)
10. The number, type, and sequence of practice items on the next version of DIBELS are sufficient to ensure that students understand the tasks	26	5.22 (.49)
12. The new administration procedures and scoring rules for the next version of DIBELS do not cause any unexpected difficulties during testing	26	5.04 (.57)
14. The next version of DIBELS scores sheets makes it easy to track items, and record a final score	25	4.92 (.64)
16. The scores obtained from the next version of DIBELS are accurate indicators of students' development of critical early literacy skills	16	4.63 (.81)
18. The opportunity to indicate Response Patterns for the next version of DIBELS measures provides valuable information in addition to the scores	23	4.56 (.68)
20. The Quality of Response rating for the retell component of ORF provides valuable information in addition to the ORF/retell score	24	4.48 (1.63)

Note. Ratings were on a 6 point scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, 6 = Strongly Agree.

Table 52. *DIBELS BGS Usability Questionnaire Ratings*

Item	N	Mean Rating (SD)
21. The Quality of Response rating for WUF-R provides valuable information in addition to the WUF-R score	23	4.53 (.75)
22. Overall, data obtained from the next version of DIBELS are helpful for goal setting	91	4.66 (.99)
24. Overall, data obtained from the next version of DIBELS are helpful for making instructional grouping decisions	89	4.81 (.89)
26. Overall, data obtained from the next version of DIBELS are helpful for identifying instructional targets	86	4.76 (.97)
28. The next version of DIBELS benchmark booklets are useful for efficient test administration and management of student scores	26	5.06 (.77)
29. The new formatting features for the next version of DIBELS materials do not cause any unexpected difficulties during testing	25	5.05 (.64)
32. I would suggest use of the next version of DIBELS to others	84	4.62 (1.10)

Note. Ratings were on a 6 point scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, 6 = Strongly Agree.

Table 53. Open-Ended DIBELS BGS Usability Questionnaire Items and Themes

Item	Themes ¹
5. Please comment on the administration and scoring rules of individual measures (n = 18 from the teacher & assessor and assessor only respondent groups).	FSF: Easy to administer (4) LNF: Easy (2) NWF: Easy to administer/follow (3) WUF-R: Time consuming/too long (4); concerns about scoring (4) DORF/Retell: Concerns about scoring/reliability (5) Daze: Easy to administer (2); time consuming/clumsy to score (2)
7. Please comment on the organization of materials for individual measures (n = 15 from the teacher & assessor and assessor only respondent groups).	DORF/Retell: well organized/easy to use (4); same page/flipping pages (3); time consuming/too many pages/long passages (3) WUF-R: time consuming/too lengthy (2)
9. Please comment on the wording of directions for individual measures (n = 12 from the teacher & assessor and assessor only respondent groups).	WUF-R: clear/understandable wording (2) DORF/Retell: clear/understandable (2)
11. Please comment on the practice items on individual measures (n = 7 from the teacher & assessor and assessor only respondent groups).	No themes identified
13. Please comment on any unexpected difficulties caused by the new administration procedures and scoring rules for individual measures (n = 8 from the teacher & assessor and assessor only respondent groups).	WUF-R: some scoring-related challenges (2)
15. Please comment on the score sheets of individual measures (n = 9 from the teacher & assessor and assessor only respondent groups).	DORF/Retell: page flipping, need to have scoring on same page (4)
17. Please comment on the accuracy of individual measures as indicators of specific literacy skills (n = 6 from the teacher & assessor respondent group).	DORF/Retell: concern about it's adequacy as an indicator of comprehension (4); reliability/consistency of scoring (2)
19. Please comment on the usefulness of the Response Patterns data for individual measures (n = 10 from the teacher & assessor and assessor only respondent groups).	DORF/Retell: takes too much time/paper (2)
23. Please comment on the usefulness of individual measures for goal setting (n = 51 from the teacher & assessor and teacher only respondent groups).	FSF: very useful/helpful (5) LNF: Useful/helpful (very) (7); Not useful/helpful (4) PSF: Useful/helpful (very) (7); don't know how to use (2) NWF: Useful/helpful (very) (8); somewhat/fairly useful (4) WUF-R: not useful/not sure how to use/difficult (7); useful/very good (6) DORF/Retell: Useful/helpful (very) (18); DORF yes, but Retell no (2); somewhat useful (2) Daze: Useful/helpful (very) (8); not used/not useful (3); somewhat (2)

Item	Themes ¹
25. Please comment on the usefulness of individual measures for making grouping decisions (n = 38 from the teacher & assessor and teacher only respondent groups).	FSF: useful/very useful (5) LNF: useful/very useful or helpful = 5; not helpful (2) PSF: useful/very useful or helpful (6); not helpful (2) NWF: useful/very useful or helpful (12); not helpful (2); somewhat/fairly (3) WUF-R: useful/very useful or helpful (6); not helpful (3); not used (3) DORF/Retell: useful/very useful or helpful (22); not helpful (2) Daze: useful/very useful or helpful (6); don't use or not helpful (4)
27. Please comment on the usefulness of individual measures for identifying instructional targets (n = 36 from the teacher & assessor and teacher only respondent groups)	FSF: useful/helpful (very) (5) LNF: useful/helpful (very) (5); not useful (2) PSF: useful/helpful (very) (6); not useful (2) NWF: useful/helpful (very) (7); somewhat/fairly (3); not useful (2) WUF-R: useful/helpful (very) (5); not useful (2); not used (2) DORF/Retell: useful/helpful (very) (13); useful w/other things (4); not useful (3) Daze: useful/helpful (very) (7); not useful (4)
30. Please comment on the new formatting features of individual measures (n = 7 from the teacher & assessor and assessor only respondent groups)	No themes identified by measure.
31. Please list any typos or other errors (e.g., grammar, spelling, etc.) you found in any of the next version of DIBELS materials (n = 6 from the teacher & assessor and assessor only respondent groups).	No themes identified by measure.
33. Please comment on your willingness to recommend the individual measures (n = 32 from the teacher & assessor and teacher only respondent groups).	FSF: willing/very willing (6) LNF: willing/very willing (6) PSF: willing/very willing (7) NWF: willing/very willing (10); don't know (2) WUF-R: willing/very willing (6); not willing/not recommended (6); don't know/not sure (2) DORF/Retell: willing/very willing (18); not willing/not recommended (2) Daze: willing/very willing (5); not willing/not recommended (4)
34. Please share with us any additional feedback you believe would make the next version of DIBELS easier to administer, or that would make the scores more meaningful for planning and evaluating reading instruction (n = 6 from the teacher & assessor and assessor only respondent groups).	No themes identified by measure.
63. Please comment on the consistency of scores from individual measures with other information you have about students in your class (n = 27 from the teacher only group).	FSF: yes/consistent/very/accurate (5); not consistent (2) LNF: yes/consistent/very/accurate (5); not consistent (3) PSF: yes/consistent/very/accurate (7) NWF: yes/consistent/very/accurate (9); not consistent (3); don't know (2)

Item	Themes ¹
	WUF-R: yes/consistent/very/accurate (2) DORF/Retell: yes/consistent/very/accurate (11); not consistent (6) Daze: yes/consistent/very/accurate (7)
66. Please share with us any additional feedback you believe would make the scores from the next version of DIBELS more meaningful for planning and evaluating reading instruction (n = 14 from the teacher only group).	LNF: Common font kids know needed (2) NWF: Have test occur later in year (2)

Notes. 1. Numbers in parentheses indicates the number of a response indicating the theme was made.

Discussion

The primary purpose of this study was to determine the levels of performance (i.e., benchmark goals and cut-points for risk) on individual DIBELS Next measures that would allow for an accurate prediction of a student's likely success in later reading outcomes. The secondary goal of this study was to evaluate the reliability and validity of DIBELS Next; the former through various methods of reliability testing, and the latter with an internal (the new DIBELS composite score) and an external (GRADE) criterion measure. This discussion summarizes the results of our data analyses by research question, delineates the limitations of the study, and details implications for future research.

Research Question 1: What levels of performance on DIBELS Next assessments predict a student is likely to achieve future reading goals and what level of performance on DIBELS Next assessments predict a student is unlikely to meet future reading goals?

The updated benchmark goals for the individual DIBELS measures in addition to the benchmark goals and cut points for risk for the DIBELS Composite Score provide important information about the likelihood of students achieving later reading goals. These benchmark goals and cut points for risk are found in Figure 2. The development of these goals is described in the results section of this technical report, beginning on page 150.

The benchmark goals for each grade and time of year were established so that a student scoring at or above the benchmark goal would have high odds of achieving future reading goals and performing equally well on the external criterion measure, the GRADE. Therefore, students scoring at or above the benchmark goal would have an 80% - 90% chance of achieving future reading goals on DIBELS Next measures.

The cut point for risk was established with a similar objective as the benchmark goal. For students scoring below the cut point for risk, the odds of achieving future reading goals on DIBELS Next measures are 10% - 20%, similar to the odds of achieving future reading goals on the GRADE.

Students with scores that fall between the benchmark goal and the cut point for risk have about even odds (40% - 60%) of achieving future reading goals on DIBELS measures. With additional support, the likelihood of the student making adequate progress increases.

Research Question 2: What is the reliability (test-retest, alternate form, and inter-rater) of the DIBELS Next assessments and the DIBELS Composite Score?

Reliability coefficients are consistently high across all three forms of reliability. The magnitude of the coefficients suggests that *DIBELS Next* possesses little test error and that users can have confidence in test results. With repeated assessment across multiple forms, reliability increases substantially, as noted where the estimated three-form reliability is reported (see Tables 19 - 22).

Inter-rater reliability coefficients for kindergarten through third grade ranged from .87 to .99. Third grade correlations ranged from .85 to .99 with the exception of Retell Quality of Response, which had a correlation of .70. Correlations for fourth through sixth grade ranged from .81 to .99 for all measures except Retell Quality of Response in fifth grade, which had a correlation of .69. Across all grades and measures, there is no evidence to indicate that the two rater's scores are significantly different. Correlations are strong and descriptive statistics illustrate the similarities in scores. Inter-rater reliability is high for all measures indicating that scoring directions are applied similarly across different assessors. To view the reliability of specific measures by grade, see Table 19 and 20.

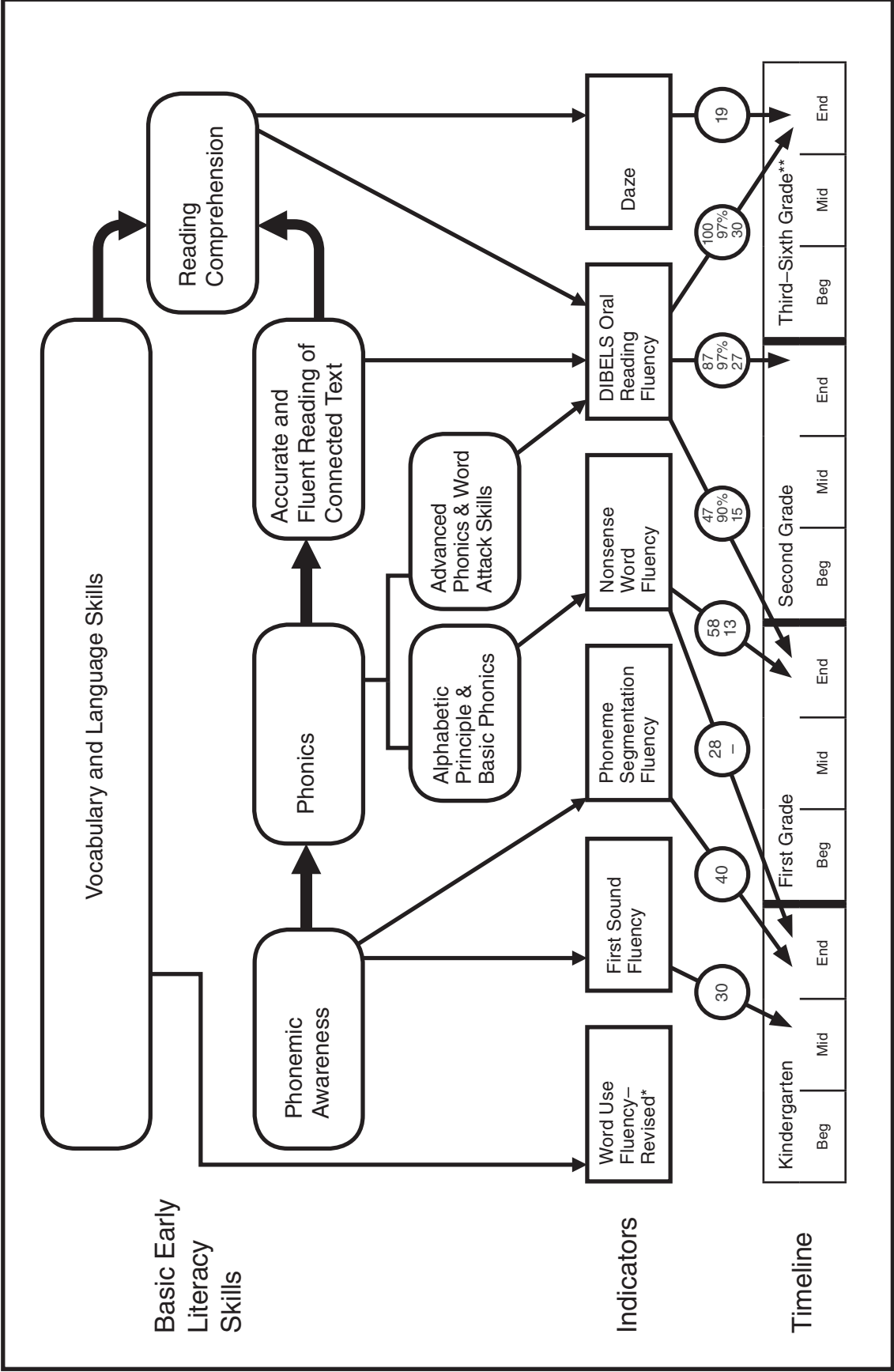
Alternate form reliability correlations for kindergarten range from .44 for PSF to .92 for NWF-WWR. Most scores on the second form, which was administered 2 weeks after the first form, are consistently higher than the scores on the first form. The higher scores may indicate that growth in skills occurred possibly aided by the effect of practice with the second form administration. Despite the slightly higher scores, moderate to strong correlations between forms were found. Correlations between the alternate forms for sixth grade are unavailable due to a small sample size (i.e., n = fewer than 15 students with scores).

Test-retest correlations for the DIBELS Next measures ranged from .67 to .97 with the exception of FSF in kindergarten, ORF Accuracy in second grade, and ORF Retell in second-, fourth-, and fifth-grade. These moderate to moderate-strong correlations indicate the stability and generally high reliability of the individual measures over time.

Research Question 3: What are the correlations for DIBELS Next assessments with each other and with the Group Reading Assessment and Diagnostic Evaluation (GRADE), a criterion measure of reading proficiency that includes comprehension?

The correlation tables for the DIBELS Next measures (Tables 23-29) indicate positive concurrent relationships and with other measures of the same skill (see Figure 26 for the model of basic early literacy skills with the DIBELS Next measures). That is, FSF in kindergarten correlates positively with PSF; LNF correlates positively with NWF CLS and with WWR to a lesser degree; NWF correlates positively with DORF Words Correct; and DORF correlates positively with Daze. The data provide evidence that the measures accurately predict each other at later time points. Negative correlations indicate the relationship between DORF errors and other DIBELS Next measures. These correlations are expected because the number of errors

should decrease with an increase in other measures of reading fluency (i.e., an increase in overall reading proficiency).



Note: For Nonsense Word Fluency, the first number is the Correct Letter Sounds goal, and the second number is the Whole Words Read Goal. For DIBELS Oral Reading Fluency, the first number is the Words Correct goal, the second number is the Accuracy goal, and the third number is the Retell goal. For example, at the end of second grade a student should be able to read 87 words with 97% accuracy and retell of 27 words relevant to the passage.

* Word Use Fluency-Revised (WUF-R) is available as an experimental measure from <http://dibels.org/>.

** Kindergarten through third grade benchmark goals are illustrated; benchmark goals for fourth through sixth grade are also available.

Figure 26. Model of Basic Early Literacy Skills, *DIBELS Next* Indicators, and Timeline

Correlation coefficients indicating the strength of the relation between the *DIBELS Next* measures and the *DIBELS* Composite score at a later time are reported in Tables 39 through 45. Overall, the predictive validity of all *DIBELS Next* 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.

In kindergarten, the predictive validity of the DCS ranges from .45 for the beginning of year DCS predicting the middle of year LNF, to .67 for the middle of year DCS predicting the end of year LNF. In first grade, the strength of the DCS score predicting the performance on individual *DIBELS* measures ranges from .71 for the beginning of year DCS predicting the middle of year NWF – WWR score, to .86 for the middle of year DCS predicting the end of year NWF – CLS score. In second grade, the validity of DCS ranges from .86, for DCS beginning of year correlated with NWF - CLS in the beginning of the year, to .88 for DCS in the middle of the year predicting ORF at the end of the year. The predictive validity of the DCS for grades three through six follow similar patterns indicating that the DCS is a strong predictor of both Daze and DORF. The relationship of DCS predicting the Daze Adjusted Score ranges from .75, for third grade beginning of year DCS predicting middle of year Daze Adjusted Score, to .79 for the same relationship between measures in fifth grade. The correlational relationship between DCS and ORF ranges from .87, for third grade middle of year DCS predicting end of year ORF, to .88 in grades four, five and six for middle of year DCS predicting end of year ORF.

Summary. The validity coefficients for kindergarten DIBELS measures based on the GRADE are moderate to moderate-strong for LNF, FSF, and NWF-CLS. The lower correlations were a result of floor effects, which are typical in beginning and middle-of-year kindergarten. We found generally lower than expected correlations for PSF (e.g., ranging from small to moderate with GRADE TT). These lower correlations may have been due to (a) ceiling effects on PSF given our higher than typically-performing sample, (b) lack of variability in GRADE scores in the sample, or (c) other factors not fully known or understood (e.g., the ability of a group-administered test to measure early literacy skills in kindergarten).

In general, correlational relationships for NWF-CLS are lower for kindergarten and increase across first and second grades and time of year. The correlations for both NWF-CLS and NWF-WWR support its concurrent and predictive validity as a measure of the alphabetic principle and beginning decoding skill.

The DORF correlations we report suggest excellent concurrent and predictive validity with the GRADE TT, as well as the VOC and COMP composites. Nearly all correlations were in the moderate-strong to strong range. Only one correlation fell slightly below the moderate-strong level and that was for DORF words correct with GRADE VOC. We also noted that overall the COMP composite showed consistently higher correlations with DORF words correct than VOC. This finding seems reasonable given the close link between the GRADE COMP task demands and the task demands for DORF.

The correlations of DORF Accuracy with GRADE TT, VOC, and COMP are moderate to strong at all three benchmark assessments and indicate very good concurrent and predictive validity for DORF Accuracy. We noted that the correlations tend to decrease as grade-level increases, which may have been due to the tendency of GRADE COMP scores to become more

uniformly distributed as student grade level increased. Overall, concurrent and predictive validity of DORF Words Correct and DORF Accuracy is well supported by the GRADE, and affirms that DORF is an excellent indicator of overall reading skill.

The correlations of Retell with GRADE TT, VOC, and COMP are moderate to moderate-strong at all three benchmark assessments indicating that retell has good concurrent and predictive validity. Correlations in first and second grades tend to decrease from middle to end-of-year assessment, but correlations in third through sixth grades tend to stay the same across time of year. The correlations of Daze Adjusted Score with GRADE TT, VOC, and COMP is moderate-strong at all three benchmark assessments. Thus, the concurrent and predictive validity of Retell and Daze Adjusted Score are also well supported by the correlations with the GRADE, and affirms that Retell and Daze are excellent indicators of reading comprehension.

The DCS correlations with the GRADE in kindergarten are stronger in the beginning and the middle of the year than the end of the year. This may be due to the different measures that comprise the DCS in kindergarten at different time points in the year. We note that predictive validity coefficients are larger than concurrent validity coefficients. The DCS is calculated from scores from different measures at beginning (FSF and LNF), middle (FSF, LNF, PSF, and NWF-CLS), and end of year (LNF, PSF, and NWF-CLS). The floor effects present in FSF and LNF are also present in beginning-of-year DCS, and thus the correlation with GRADE (with floor effects) is artificially deflated.

The DCS correlations with the GRADE in kindergarten are stronger in the beginning and the middle of the year than the end of the year. This may be due to the different measures that comprise the DCS in kindergarten at different time points in the year. We note that predictive validity coefficients are larger than concurrent validity coefficients. The DCS is calculated from

scores from different measures at beginning (FSF and LNF), middle (FSF, LNF, PSF, and NWF-CLS), and end of year (LNF, PSF, and NWF-CLS). The floor effects present in FSF and LNF are also present in beginning-of-year DCS, and thus the correlation with GRADE (with floor effects) is artificially deflated.

Research Question 4: To what extent are teachers and assessors satisfied with the measures?

Across all items on the DIBELS BGS Usability Questionnaire, high levels of agreement were found for every item regardless of respondent group. For every item requiring a rating response, responses were mostly for the "Agree," and "Strongly Agree" ratings. However, items also were rated by some respondents with varying levels of disagreement ranging from "Strongly Disagree" to "Somewhat Disagree." Feedback on open-ended items also was very positive. Overall themes included ease of administration and use; utility for goal setting, grouping decisions, and identifying instructional targets; general willingness to recommend the measures; and the belief that the measures were generally consistent with other information teachers had about their students. Many issues mentioned in the comments correspond to changes ultimately made in test materials (e.g., child-friendly font in K - 2 materials, scoring section for DORF on the first page of the passage; response patterns for DORF after the third passage resulting in fewer pages overall, etc.)

Limitations

As with all research, this study had several possible limitations. One issue was that overall, the students that participated in the Benchmark Goals Study represented a relatively high-achieving sample. For example, 73% of the third grade sample scored at or above the 40th percentile on the GRADE. This limitation was addressed by taking the high-achieving sample into consideration when setting the benchmark goals. For example, the third grade end of year benchmark goals were set so that 73% of the sample scored at or above the goal on the DIBELS Next Composite Score. Likely due to the relatively high achieving sample, ceiling effects appeared to impact some of the data collected. More specifically, the ceiling effects present in kindergarten GRADE scores may artificially deflated correlations with DIBELS measures.

Another possible limitation to consider is that data were collected in a naturalistic setting with no additional controls on the reliability of administration and scoring, beyond the training assessors received and the fact that all sites were long-term DIBELS users who are well-versed in DIBELS administration and scoring. In addition, our knowledge of specific instructional context for each site is limited, although we do know that four of the five districts have a unified systematic approach to reading instruction that is used school-side and that all districts were long-term users of DIBELS.

With respect to the DIBELS BGS Usability Questionnaire, the sample of respondents was comprised of teachers and other educational personnel who volunteered for this study. Thus, selection bias is a potential limitation. Given that these educators volunteered their participation, respondents might have been biased to respond positively. Alternatively, some respondents might have felt pressured to complete additional paperwork and responded more negatively as a result.

In addition, the questionnaire data are not a direct means of determining educator opinions about the usability of DIBELS Next. Because the data were collected via an anonymous electronic questionnaire, there is no means to follow up with respondents about why they responded in the manner in which they did. Further research could address this limitation by gathering qualitative data from focus groups..

Finally, as with all studies of this nature, it is possible that results may differ with a different sample of students or a different external criterion measure. We mediated this concern in part by using an improper linear models approach in the analysis to determine benchmark goals and cut-points for risk.

Implications for Research and Practice

Future research should include repeating this study with additional samples of students representing a wider range of student performance across all measures to allow for greater variability in the data, mediate the impact of ceiling effects, and increase the generalizability of results. Replication can also include samples of students that represent greater geographical and socio-economic diversity than the current study. A variety of different outcome (or criterion) measures may also be utilized to replicate the results of this study.

With respect to implications for practice, it should be noted that the only new measure for this study was the DIBELS Composite Score (DCS). This new measure is the best indicator of the student's overall reading proficiency. This represents a change from the DIBELS 6th Edition procedures where DIBELS ORF was used as the primary indicator of a student's reading proficiency. Additional implications are present in the changes in directions for administration of other DIBELS Next measures. For example, the changes to directions for DORF represent tighter control with regard to readability of the passages and help to minimize score variability.

For NWF, the changes to directions were designed to encourage students to say the words if they could. Additional measures, including modifications to directions, that are new for DIBELS Next were included in previous research studies.

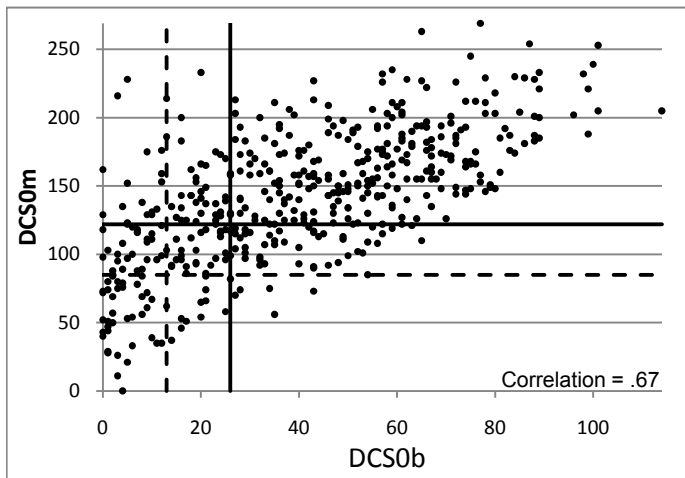
For more information about the changes to the measures, benchmark goals, or further information about their validity, please refer to Technical Reports 7 and 10 (available at www.dibels.org), and the DIBELS Next Technical Manual (2011).

Appendix A : DIBELS Next Benchmark Goals and Cut Points for Risk Analysis Detail

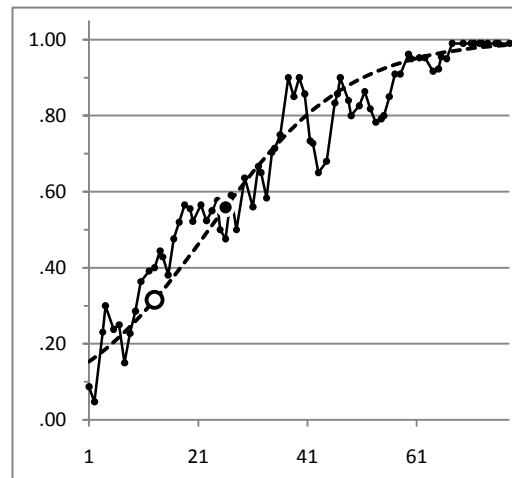
Pages.

This appendix contains a summary of the analyses that were used to support each benchmark goal and cut point for risk for all measures by grade level and time of year. This include the relationships between all the DIBELS next measures and both the internal and external criterion measures.

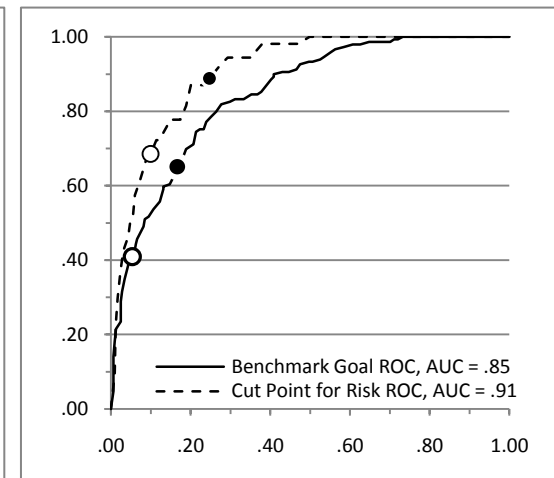
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS0b	26	13	DIBELS Composite Score, Grade K, Beginning of Year
Outcome	Criterion	DCS0m	122	85	DIBELS Composite Score, Grade K, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

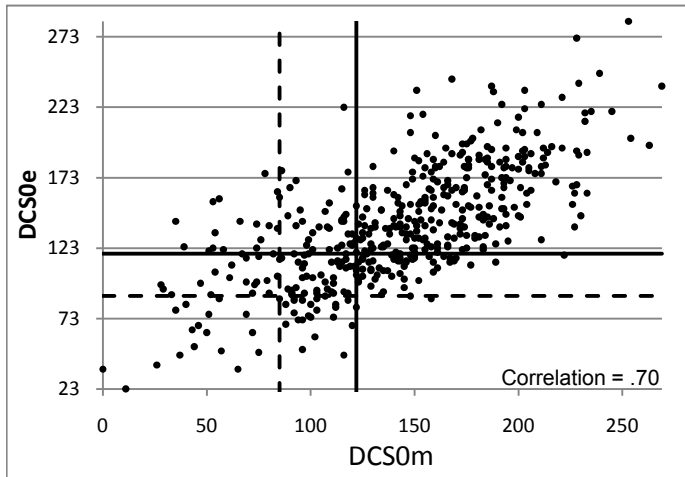
DCS0b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0m Outcome:					
At or Above Benchmark	17	36	265	318	68%
Below Benchmark	24	25	46	95	20%
Well Below Benchmark	37	11	6	54	12%
Marginal Total	78	72	317	467	
Marginal Percent	17%	15%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	22%	50%	84%		

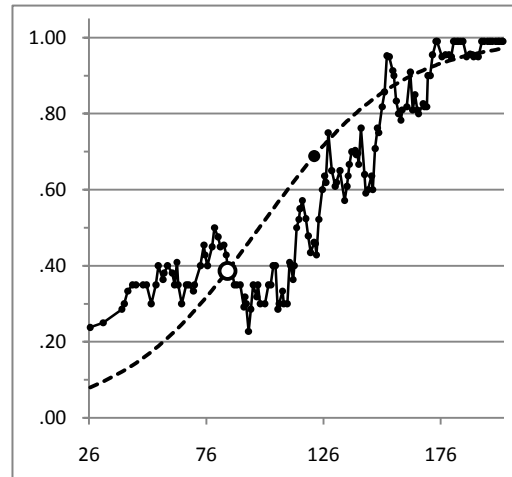
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	265	301	311	372
False Negative	52	88	6	17
True Positive	97	61	48	37
False Positive	53	17	102	41
Sensitivity	.65	.41	.89	.69
specificity	.83	.95	.75	.90
Negative Predictive Power	.84	.77	.98	.96
Positive Predictive Power	.65	.78	.32	.47
Accurate Classification	.78	.78	.77	.88
Kappa	.48	.41	.36	.49

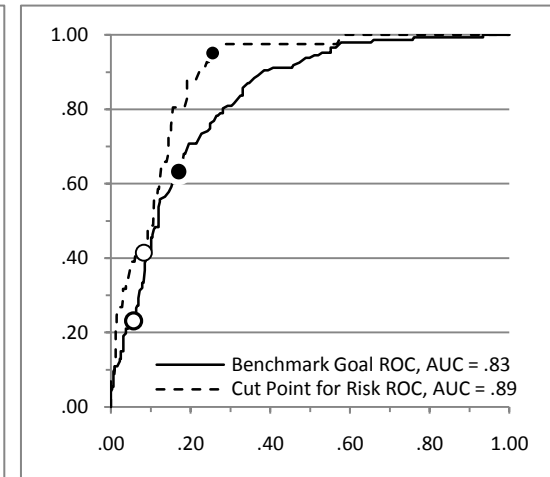
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS0m	122	85	DIBELS Composite Score, Grade K, Middle of Year
Outcome	Criterion	DCS0e	119	89	DIBELS Composite Score, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

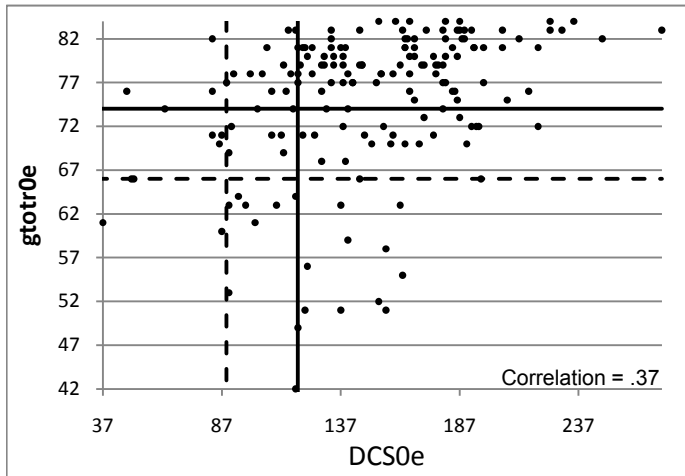
DCS0m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0e Outcome:					
At or Above Benchmark	18	36	263	317	68%
Below Benchmark	17	37	52	106	23%
Well Below Benchmark	17	22	2	41	9%
Marginal Total	52	95	317	464	
Marginal Percent	11%	20%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	35%	38%	83%		

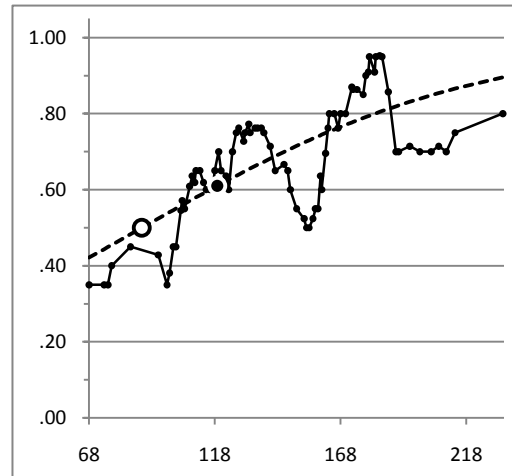
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	263	299	315	388
False Negative	54	113	2	24
True Positive	93	34	39	17
False Positive	54	18	108	35
Sensitivity	.63	.23	.95	.41
specificity	.83	.94	.74	.92
Negative Predictive Power	.83	.73	.99	.94
Positive Predictive Power	.63	.65	.27	.33
Accurate Classification	.77	.72	.76	.87
Kappa	.46	.21	.32	.30

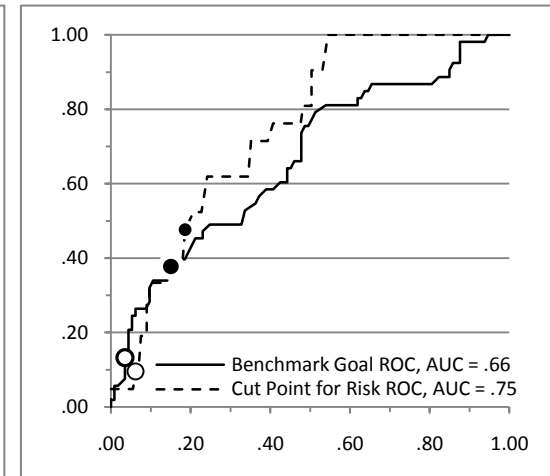
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS0e	119	89	DIBELS Composite Score, Grade K, End of Year
Outcome	Criterion	gtotr0e	74	66	GRADE Total Test, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

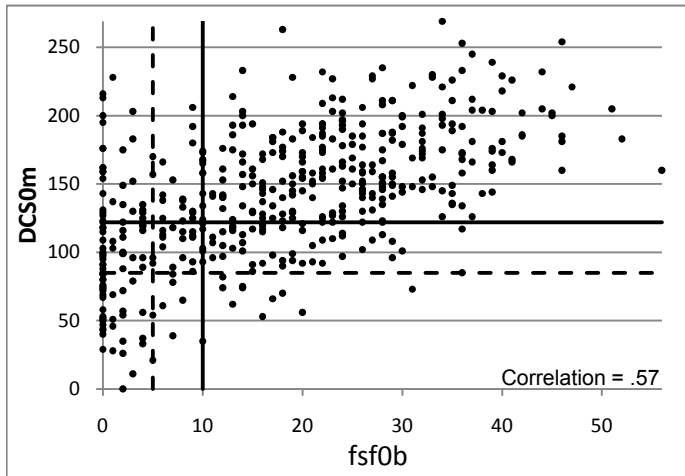
DCS0e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr0e Outcome:					
At or Above Benchmark	4	13	96	113	68%
Below Benchmark	5	5	22	32	19%
Well Below Benchmark	2	8	11	21	13%
Marginal Total	11	26	129	166	
Marginal Percent	7%	16%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	36%	50%	74%		

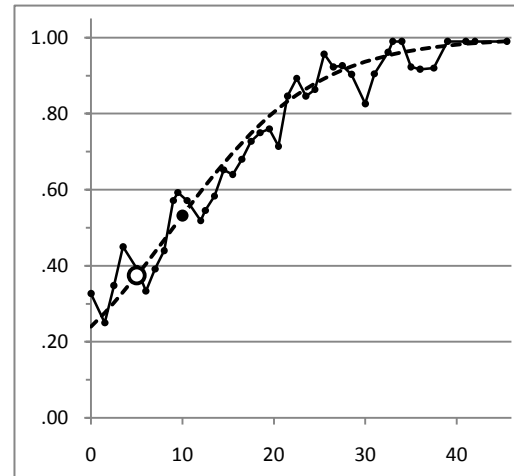
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	96	109	118	136
False Negative	33	46	11	19
True Positive	20	7	10	2
False Positive	17	4	27	9
Sensitivity	.38	.13	.48	.10
specificity	.85	.96	.81	.94
Negative Predictive Power	.74	.70	.91	.88
Positive Predictive Power	.54	.64	.27	.18
Accurate Classification	.70	.70	.77	.83
Kappa	.25	.12	.22	.04

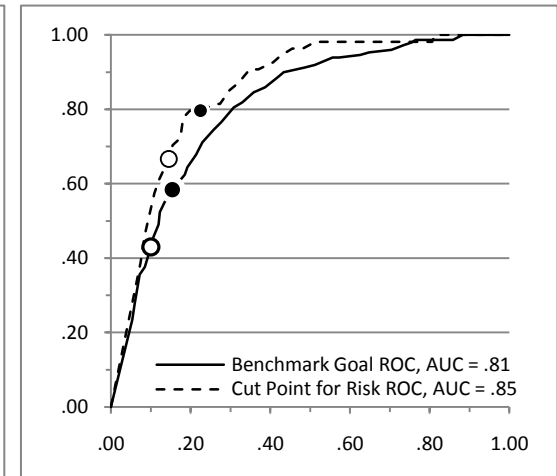
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	fsf0b	10	5	FSF, Grade K, Beginning of Year
Outcome	Criterion	DCS0m	122	85	DIBELS Composite Score, Grade K, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

fsf0b Screening Decision:

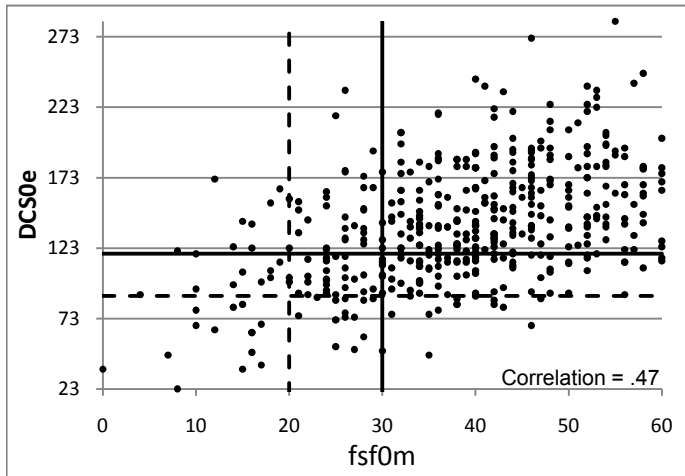
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0m Outcome:					
At or Above Benchmark	32	17	269	318	68%
Below Benchmark	28	16	51	95	20%
Well Below Benchmark	36	7	11	54	12%
Marginal Total	96	40	331	467	
Marginal Percent	21%	9%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	33%	43%	81%		

At or Above Benchmark outcome

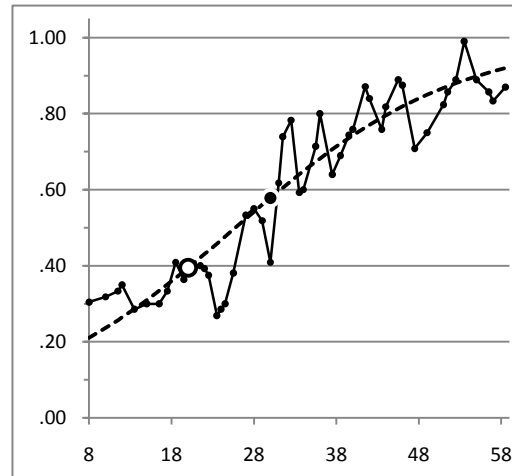
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	269	286	320	353
False Negative	62	85	11	18
True Positive	87	64	43	36
False Positive	49	32	93	60
Sensitivity	.58	.43	.80	.67
specificity	.85	.90	.77	.85
Negative Predictive Power	.81	.77	.97	.95
Positive Predictive Power	.64	.67	.32	.38
Accurate Classification	.76	.75	.78	.83
Kappa	.44	.36	.34	.39

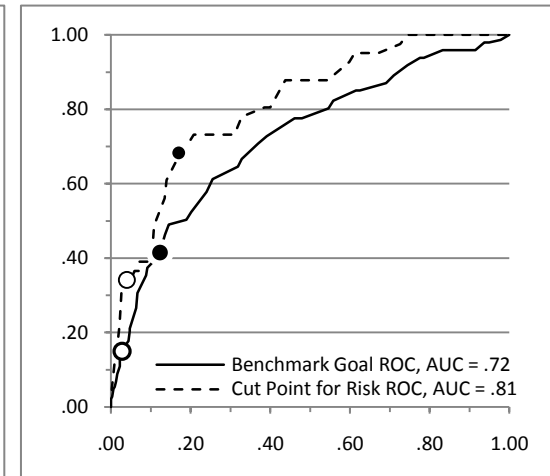
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	fsf0m	30	20	FSF, Grade K, Middle of Year
Outcome	Criterion	DCS0e	119	89	DIBELS Composite Score, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

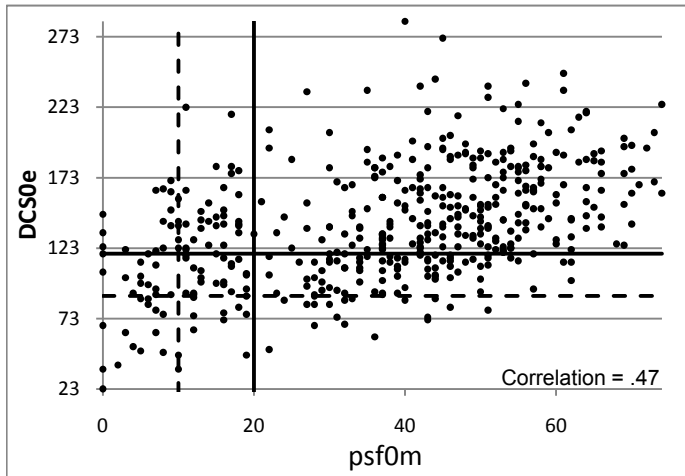
fsf0m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0e Outcome:					
At or Above Benchmark	9	30	278	317	68%
Below Benchmark	8	25	73	106	23%
Well Below Benchmark	14	14	13	41	9%
Marginal Total	31	69	364	464	
Marginal Percent	7%	15%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	29%	43%	76%		

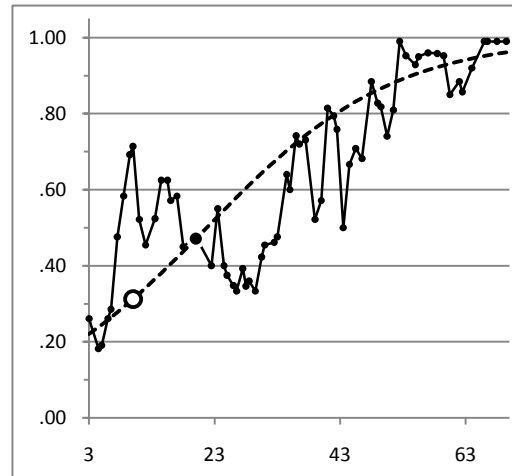
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	278	308	351	406
False Negative	86	125	13	27
True Positive	61	22	28	14
False Positive	39	9	72	17
Sensitivity	.41	.15	.68	.34
specificity	.88	.97	.83	.96
Negative Predictive Power	.76	.71	.96	.94
Positive Predictive Power	.61	.71	.28	.45
Accurate Classification	.73	.71	.82	.91
Kappa	.32	.15	.31	.34

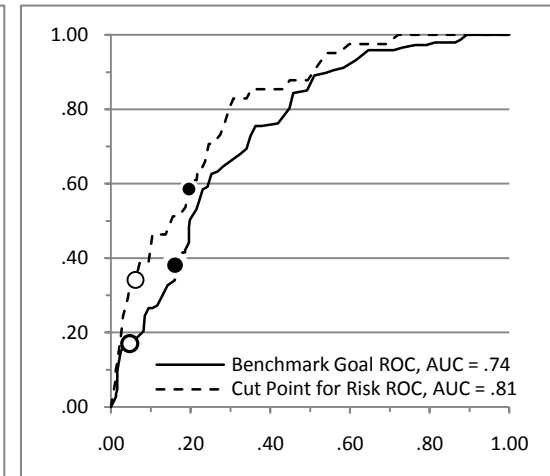
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	psf0m	20	10	PSF, Grade K, Middle of Year
Outcome	Criterion	DCS0e	119	89	DIBELS Composite Score, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

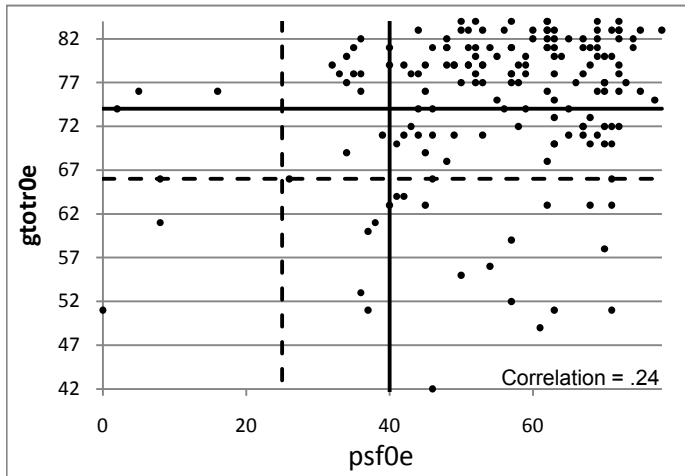
psf0m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0e Outcome:					
At or Above Benchmark	15	36	266	317	68%
Below Benchmark	11	21	74	106	23%
Well Below Benchmark	14	10	17	41	9%
Marginal Total	40	67	357	464	
Marginal Percent	9%	14%	77%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	38%	54%	75%		

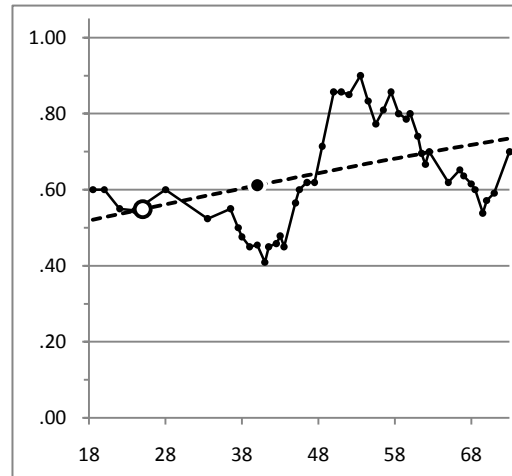
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	266	302	340	397
False Negative	91	122	17	27
True Positive	56	25	24	14
False Positive	51	15	83	26
Sensitivity	.38	.17	.59	.34
specificity	.84	.95	.80	.94
Negative Predictive Power	.75	.71	.95	.94
Positive Predictive Power	.52	.63	.22	.35
Accurate Classification	.69	.70	.78	.89
Kappa	.24	.15	.23	.28

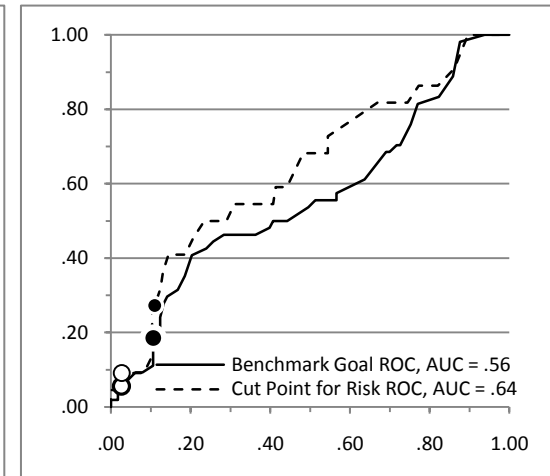
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	psf0e	40	25	PSF, Grade K, End of Year
Outcome	Criterion	gtotr0e	74	66	GRADE Total Test, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

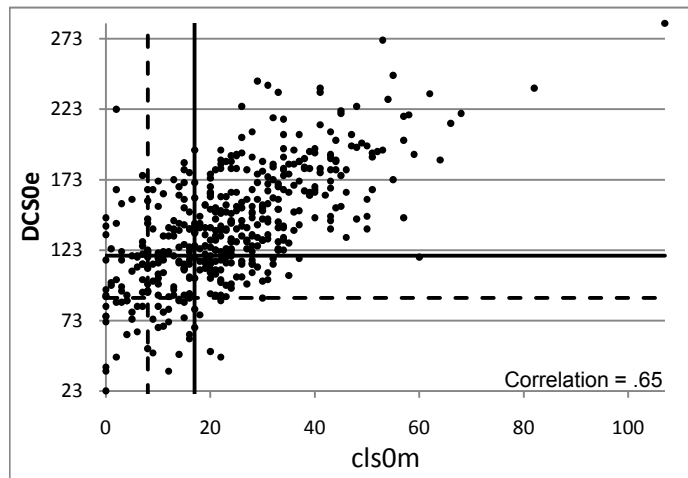
psf0e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr0e Outcome:					
At or Above Benchmark	3	9	101	113	68%
Below Benchmark	1	3	28	32	19%
Well Below Benchmark	2	4	16	22	13%
Marginal Total	6	16	145	167	
Marginal Percent	4%	10%	87%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	50%	56%	70%		

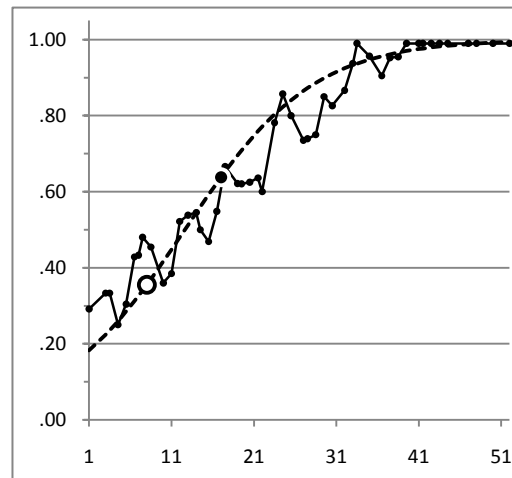
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	101	110	129	141
False Negative	44	51	16	20
True Positive	10	3	6	2
False Positive	12	3	16	4
Sensitivity	.19	.06	.27	.09
specificity	.89	.97	.89	.97
Negative Predictive Power	.70	.68	.89	.88
Positive Predictive Power	.45	.50	.27	.33
Accurate Classification	.66	.68	.81	.86
Kappa	.09	.04	.16	.09

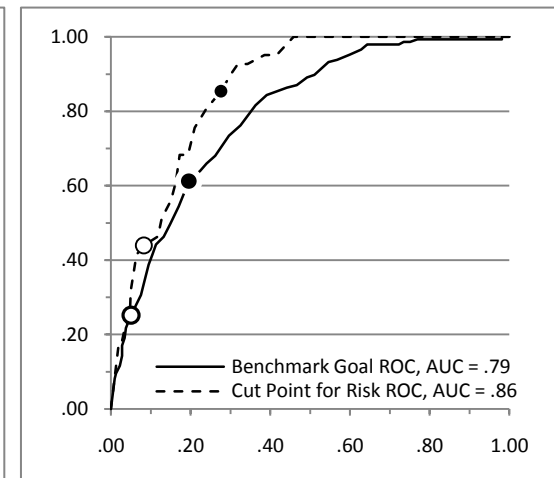
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls0m	17	8	NWF Correct Letter Sounds, Grade K, Middle of Year
Outcome	Criterion	DCS0e	119	89	DIBELS Composite Score, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

cls0m Screening Decision:

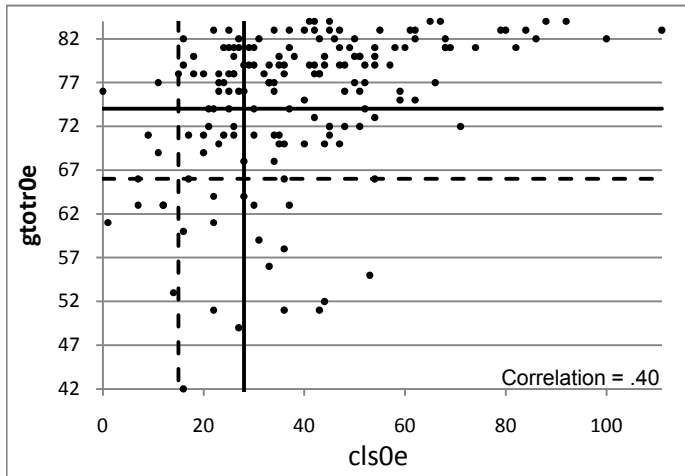
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS0e Outcome:					
At or Above Benchmark	16	46	255	317	68%
Below Benchmark	19	36	51	106	23%
Well Below Benchmark	18	17	6	41	9%
Marginal Total	53	99	312	464	
Marginal Percent	11%	21%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	30%	46%	82%		

At or Above Benchmark outcome

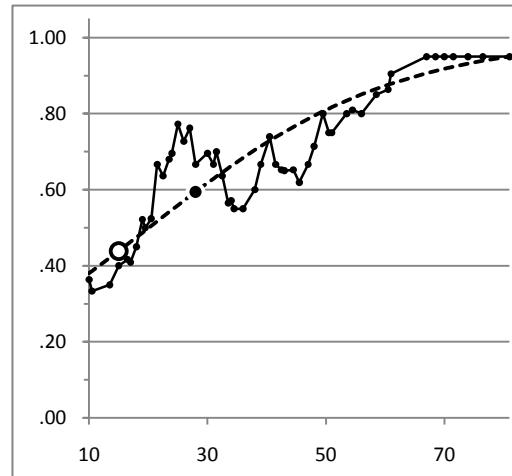
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	255	301	306	388
False Negative	57	110	6	23
True Positive	90	37	35	18
False Positive	62	16	117	35
Sensitivity	.61	.25	.85	.44
specificity	.80	.95	.72	.92
Negative Predictive Power	.82	.73	.98	.94
Positive Predictive Power	.59	.70	.23	.34
Accurate Classification	.74	.73	.73	.88
Kappa	.41	.24	.26	.31

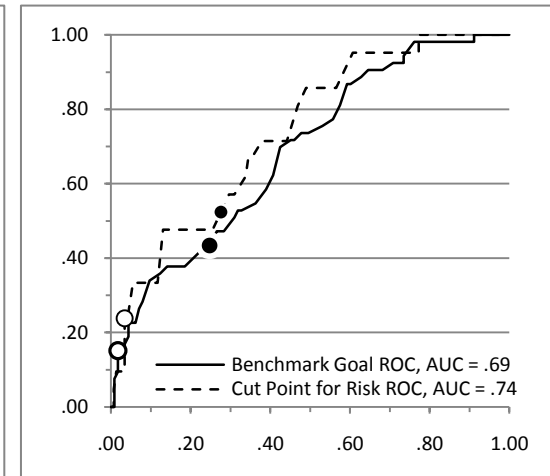
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls0e	28	15	NWF Correct Letter Sounds, Grade K, End of Year
Outcome	Criterion	gtotr0e	74	66	GRADE Total Test, Grade K, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

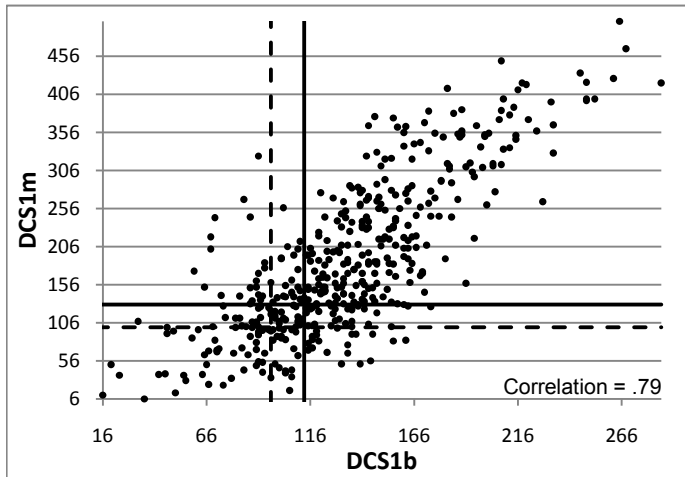
cls0e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr0e Outcome:					
At or Above Benchmark	2	26	85	113	68%
Below Benchmark	3	9	20	32	19%
Well Below Benchmark	5	6	10	21	13%
Marginal Total	10	41	115	166	
Marginal Percent	6%	25%	69%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	63%	74%		

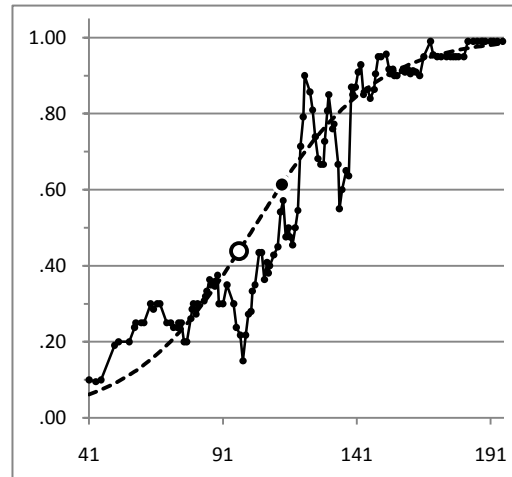
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	85	111	105	140
False Negative	30	45	10	16
True Positive	23	8	11	5
False Positive	28	2	40	5
Sensitivity	.43	.15	.52	.24
specificity	.75	.98	.72	.97
Negative Predictive Power	.74	.71	.91	.90
Positive Predictive Power	.45	.80	.22	.50
Accurate Classification	.65	.72	.70	.87
Kappa	.19	.17	.15	.26

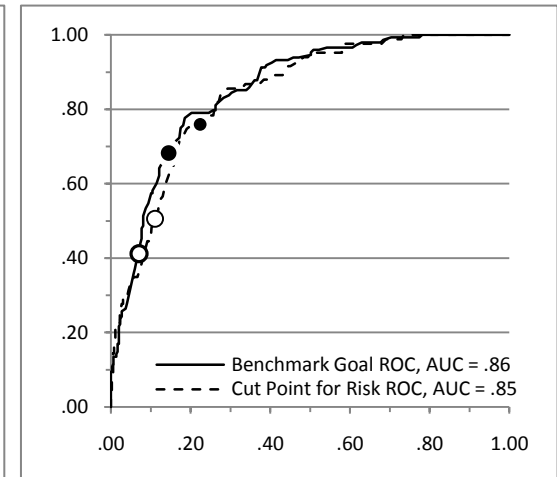
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS1b	113	97	DIBELS Composite Score, Grade 1, Beginning of Year
Outcome	Criterion	DCS1m	130	100	DIBELS Composite Score, Grade 1, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

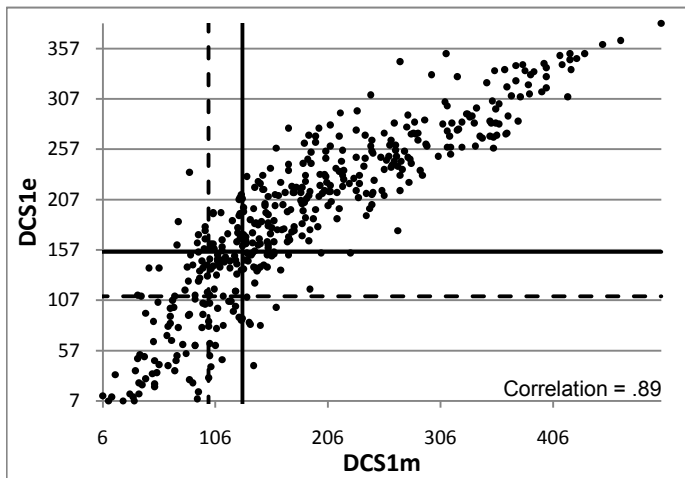
DCS1b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1m Outcome:					
At or Above Benchmark	21	22	254	297	67%
Below Benchmark	19	19	27	65	15%
Well Below Benchmark	42	21	20	83	19%
Marginal Total	82	62	301	445	
Marginal Percent	18%	14%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	35%	84%		

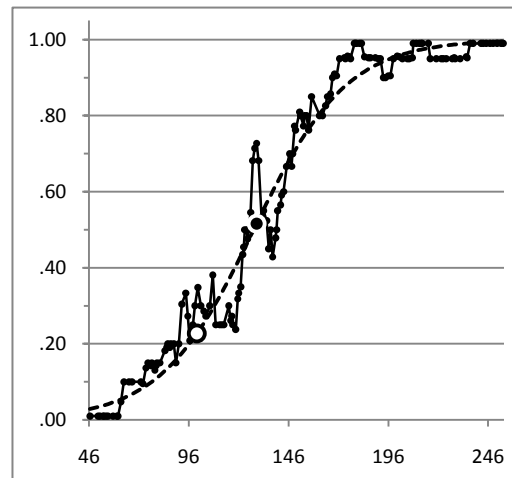
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	254	276	281	322
False Negative	47	87	20	41
True Positive	101	61	63	42
False Positive	43	21	81	40
Sensitivity	.68	.41	.76	.51
specificity	.86	.93	.78	.89
Negative Predictive Power	.84	.76	.93	.89
Positive Predictive Power	.70	.74	.44	.51
Accurate Classification	.80	.76	.77	.82
Kappa	.54	.38	.42	.40

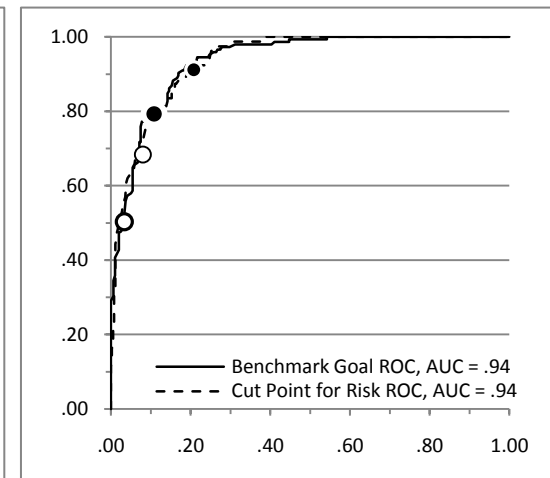
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS1m	130	100	DIBELS Composite Score, Grade 1, Middle of Year
Outcome	Criterion	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

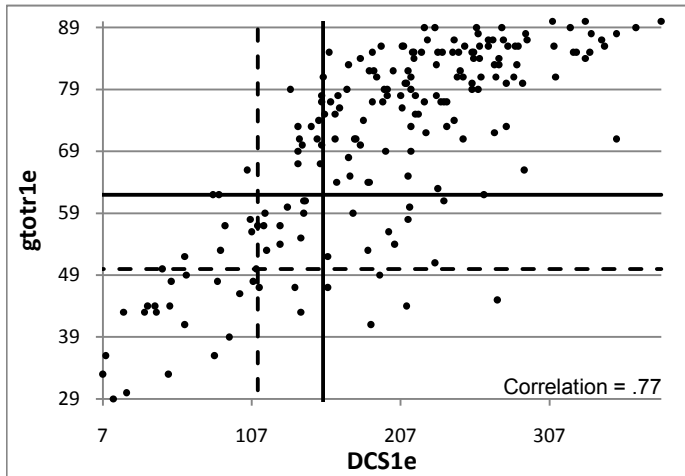
DCS1m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1e Outcome:					
At or Above Benchmark	10	22	263	295	67%
Below Benchmark	19	24	23	66	15%
Well Below Benchmark	54	18	7	79	18%
Marginal Total	83	64	293	440	
Marginal Percent	19%	15%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	34%	90%		

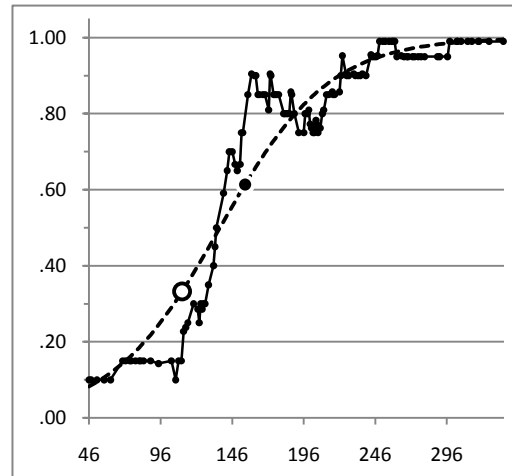
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	263	285	286	332
False Negative	30	72	7	25
True Positive	115	73	72	54
False Positive	32	10	75	29
Sensitivity	.79	.50	.91	.68
specificity	.89	.97	.79	.92
Negative Predictive Power	.90	.80	.98	.93
Positive Predictive Power	.78	.88	.49	.65
Accurate Classification	.86	.81	.81	.88
Kappa	.68	.53	.53	.59

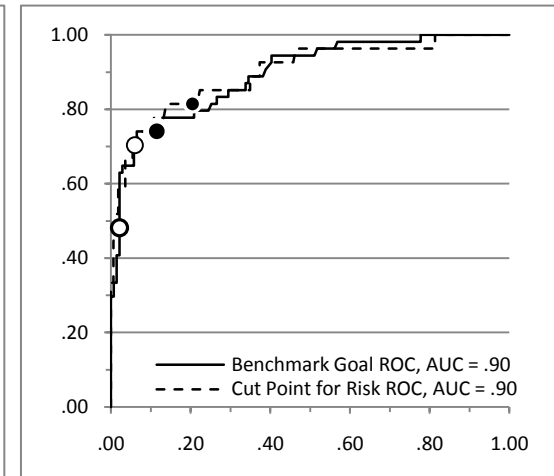
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

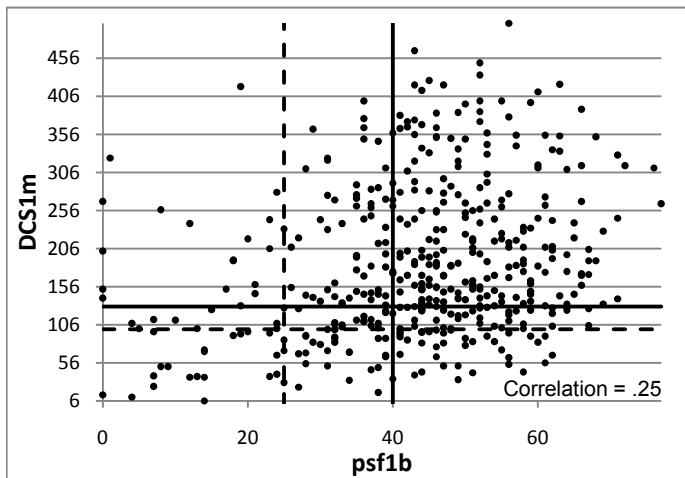
DCS1e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	3	13	123	139	72%
Below Benchmark	7	11	9	27	14%
Well Below Benchmark	19	3	5	27	14%
Marginal Total	29	27	137	193	
Marginal Percent	15%	14%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	48%	90%		

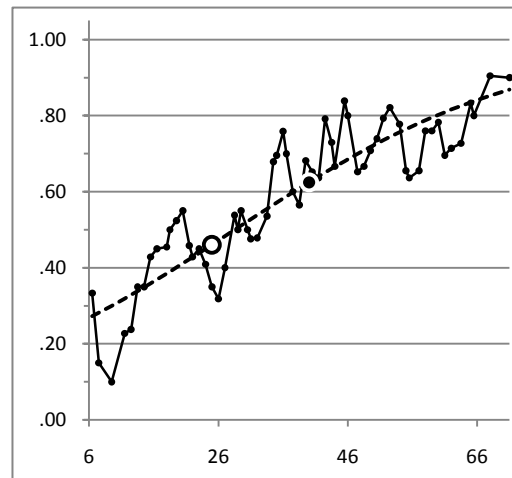
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	136	132	156
False Negative	14	28	5	8
True Positive	40	26	22	19
False Positive	16	3	34	10
Sensitivity	.74	.48	.81	.70
specificity	.88	.98	.80	.94
Negative Predictive Power	.90	.83	.96	.95
Positive Predictive Power	.71	.90	.39	.66
Accurate Classification	.84	.84	.80	.91
Kappa	.62	.54	.42	.62

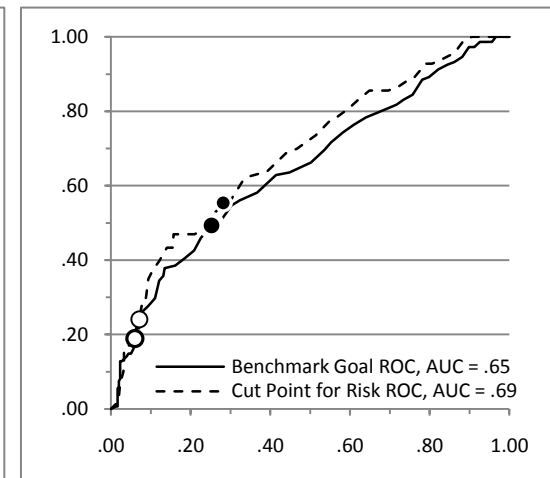
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	psf1b	40	25	PSF, Grade 1, Beginning of Year
Outcome	Criterion	DCS1m	130	100	DIBELS Composite Score, Grade 1, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

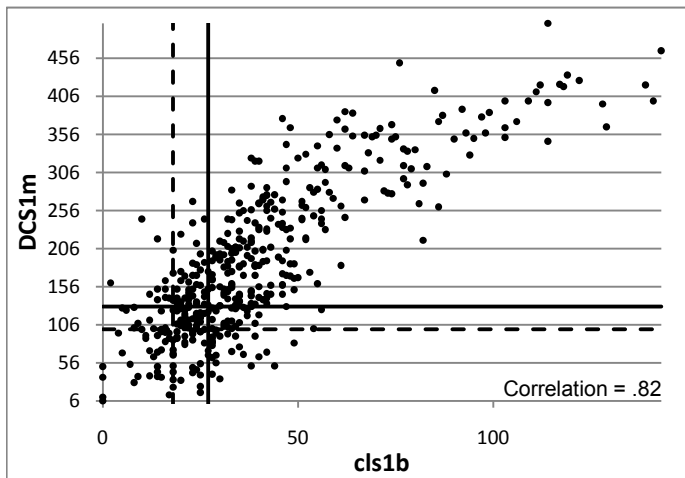
psf1b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1m Outcome:					
At or Above Benchmark	18	57	222	297	67%
Below Benchmark	8	19	38	65	15%
Well Below Benchmark	20	26	37	83	19%
Marginal Total	46	102	297	445	
Marginal Percent	10%	23%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	39%	56%	75%		

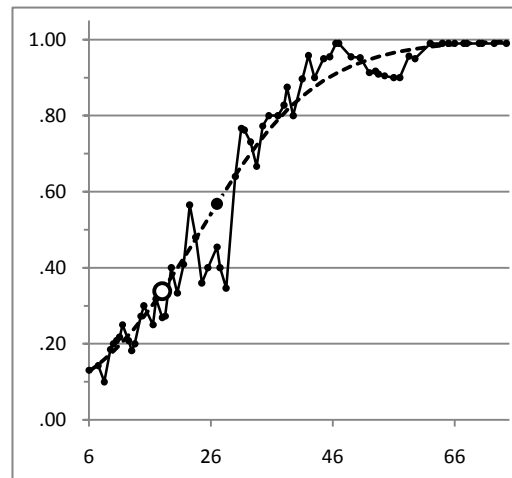
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	222	279	260	336
False Negative	75	120	37	63
True Positive	73	28	46	20
False Positive	75	18	102	26
Sensitivity	.49	.19	.55	.24
specificity	.75	.94	.72	.93
Negative Predictive Power	.75	.70	.88	.84
Positive Predictive Power	.49	.61	.31	.43
Accurate Classification	.66	.69	.69	.80
Kappa	.24	.16	.21	.20

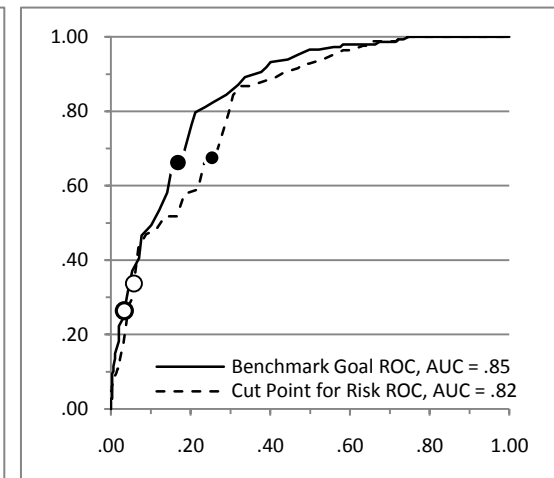
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls1b	27	18	NWF Correct Letter Sounds, Grade 1, Beginning of Year
Outcome	Criterion	DCS1m	130	100	DIBELS Composite Score, Grade 1, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

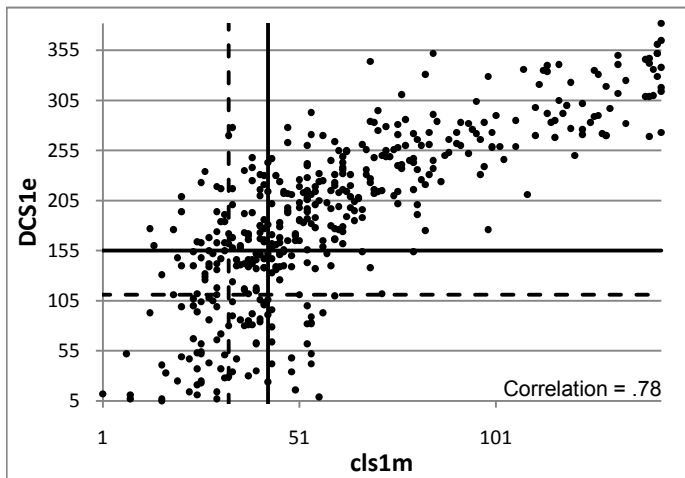
cls1b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1m Outcome:					
At or Above Benchmark	10	40	247	297	67%
Below Benchmark	11	31	23	65	15%
Well Below Benchmark	28	28	27	83	19%
Marginal Total	49	99	297	445	
Marginal Percent	11%	22%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	40%	83%		

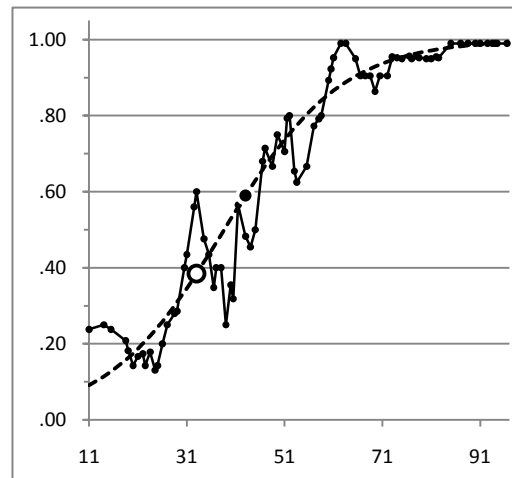
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	247	287	270	341
False Negative	50	109	27	55
True Positive	98	39	56	28
False Positive	50	10	92	21
Sensitivity	.66	.26	.67	.34
specificity	.83	.97	.75	.94
Negative Predictive Power	.83	.72	.91	.86
Positive Predictive Power	.66	.80	.38	.57
Accurate Classification	.78	.73	.73	.83
Kappa	.49	.28	.32	.33

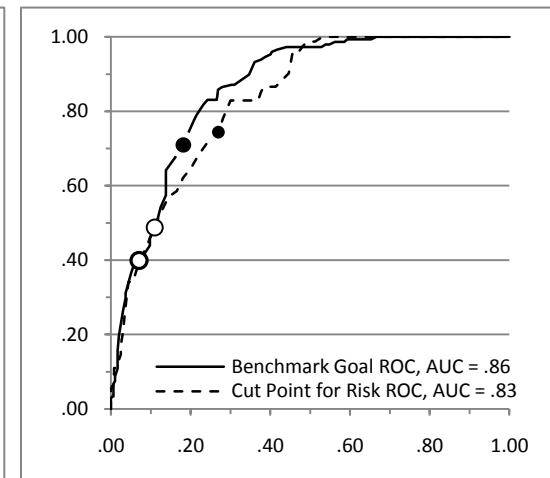
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls1m	43	33	NWF Correct Letter Sounds, Grade 1, Middle of Year
Outcome	Criterion	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

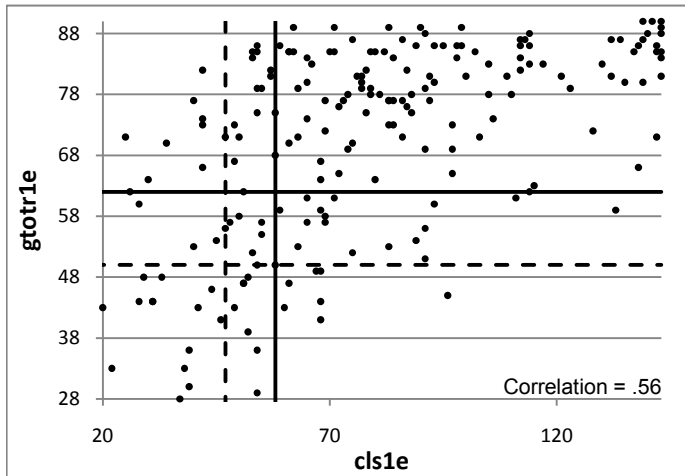
cls1m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1e Outcome:					
At or Above Benchmark	21	33	243	297	67%
Below Benchmark	19	25	22	66	15%
Well Below Benchmark	40	21	21	82	18%
Marginal Total	80	79	286	445	
Marginal Percent	18%	18%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	42%	85%		

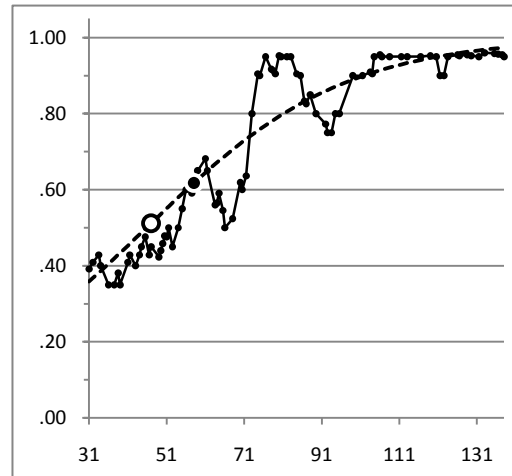
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	243	276	265	323
False Negative	43	89	21	42
True Positive	105	59	61	40
False Positive	54	21	98	40
Sensitivity	.71	.40	.74	.49
specificity	.82	.93	.73	.89
Negative Predictive Power	.85	.76	.93	.88
Positive Predictive Power	.66	.74	.38	.50
Accurate Classification	.78	.75	.73	.82
Kappa	.52	.37	.35	.38

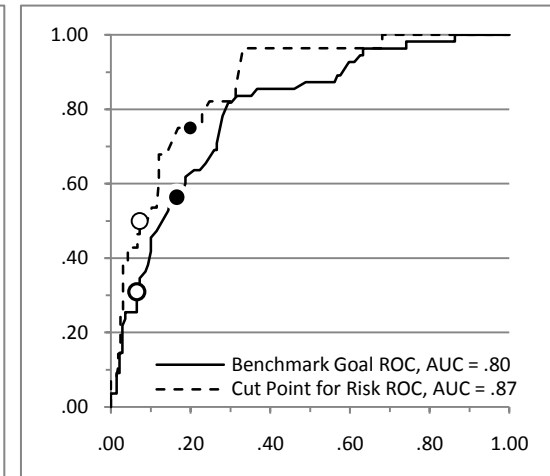
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls1e	58	47	NWF Correct Letter Sounds, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

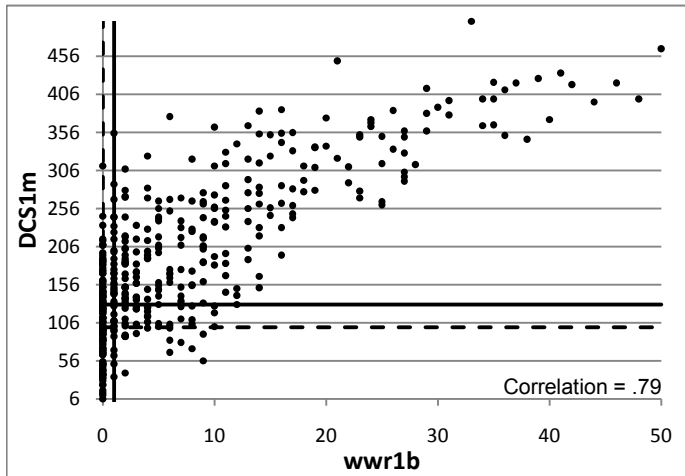
cls1e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	9	14	116	139	72%
Below Benchmark	3	7	17	27	14%
Well Below Benchmark	14	7	7	28	14%
Marginal Total	26	28	140	194	
Marginal Percent	13%	14%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	35%	50%	83%		

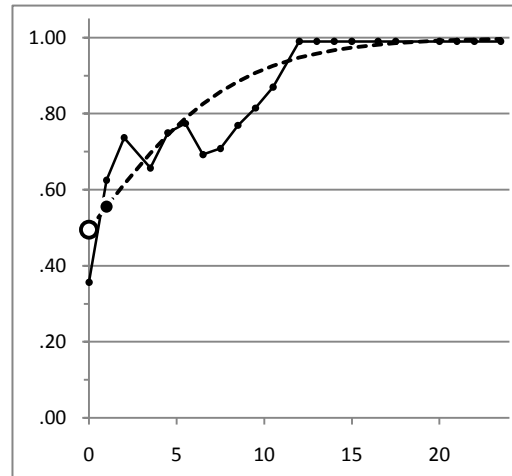
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	116	130	133	154
False Negative	24	38	7	14
True Positive	31	17	21	14
False Positive	23	9	33	12
Sensitivity	.56	.31	.75	.50
specificity	.83	.94	.80	.93
Negative Predictive Power	.83	.77	.95	.92
Positive Predictive Power	.57	.65	.39	.54
Accurate Classification	.76	.76	.79	.87
Kappa	.40	.29	.40	.44

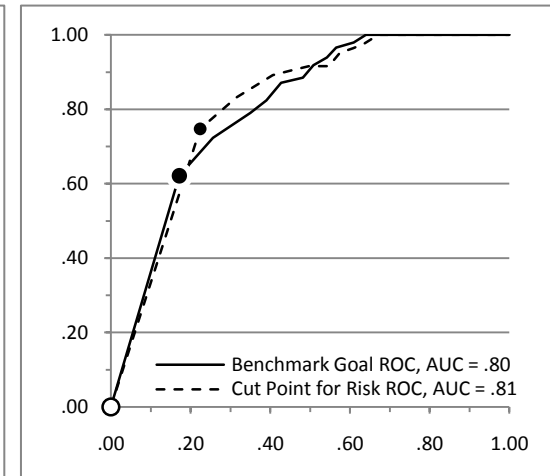
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	wwr1b	1	0	NWF Whole Words Read, Grade 1, Beginning of Year
Outcome	Criterion	DCS1m	130	100	DIBELS Composite Score, Grade 1, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

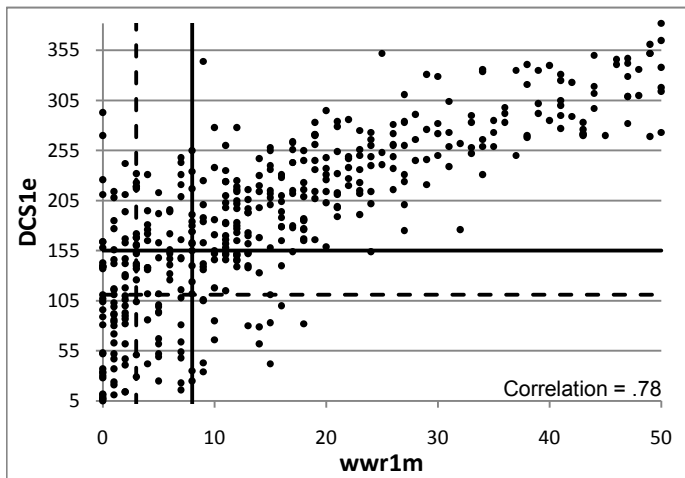
wwr1b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1m Outcome:					
At or Above Benchmark	0	51	246	297	67%
Below Benchmark	0	30	35	65	15%
Well Below Benchmark	0	62	21	83	19%
Marginal Total	0	143	302	445	
Marginal Percent	0%	32%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	36%	81%		

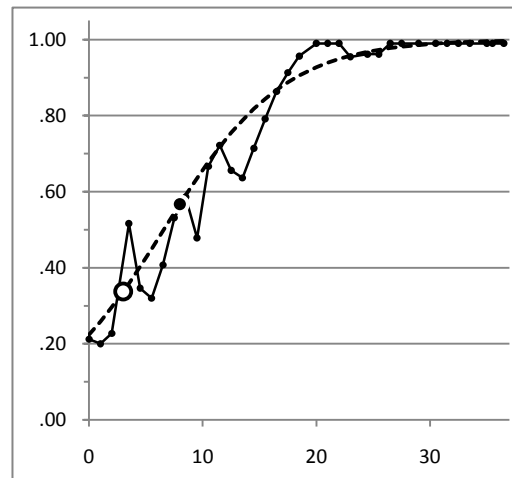
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	246	297	281	362
False Negative	56	148	21	83
True Positive	92	0	62	0
False Positive	51	0	81	0
Sensitivity	.62	.00	.75	.00
specificity	.83	1.00	.78	1.00
Negative Predictive Power	.81	.67	.93	.81
Positive Predictive Power	.64	--	.43	--
Accurate Classification	.76	.67	.77	.81
Kappa	.45	.00	.41	.00

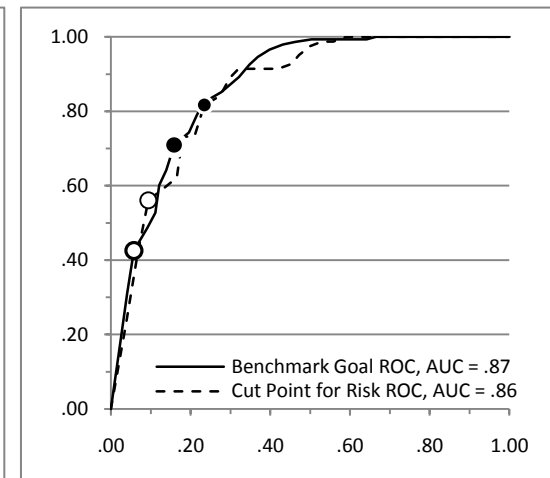
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	wwr1m	8	3	NWF Whole Words Read, Grade 1, Middle of Year
Outcome	Criterion	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

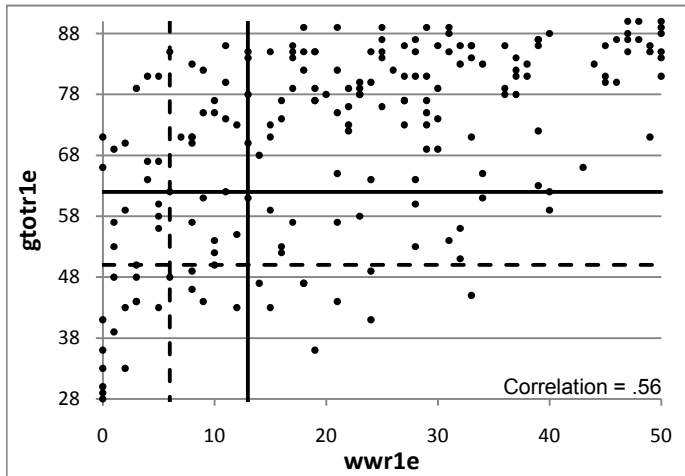
wwr1m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1e Outcome:					
At or Above Benchmark	17	30	249	296	67%
Below Benchmark	17	21	28	66	15%
Well Below Benchmark	46	21	15	82	18%
Marginal Total	80	72	292	444	
Marginal Percent	18%	16%	66%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	21%	42%	85%		

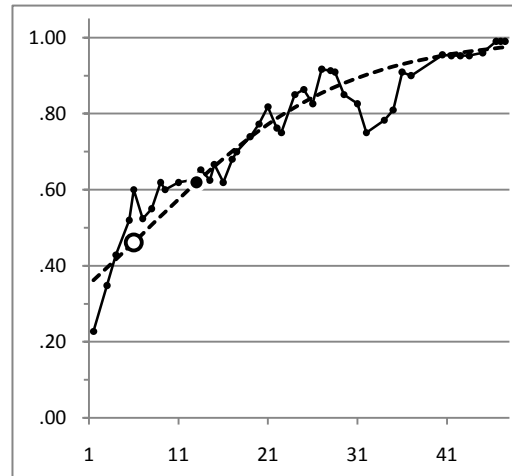
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	249	279	277	328
False Negative	43	85	15	36
True Positive	105	63	67	46
False Positive	47	17	85	34
Sensitivity	.71	.43	.82	.56
specificity	.84	.94	.77	.91
Negative Predictive Power	.85	.77	.95	.90
Positive Predictive Power	.69	.79	.44	.58
Accurate Classification	.80	.77	.77	.84
Kappa	.55	.42	.44	.47

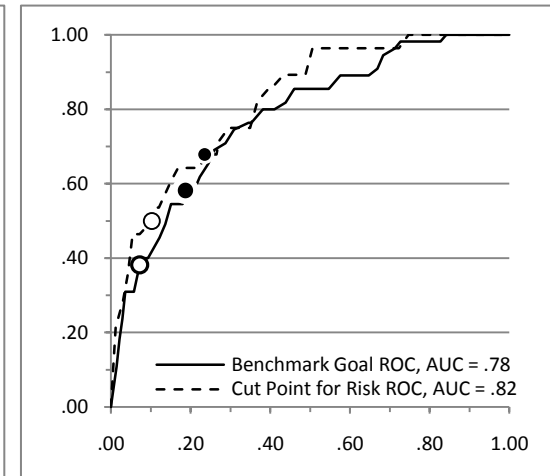
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	wwr1e	13	6	NWF Whole Words Read, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

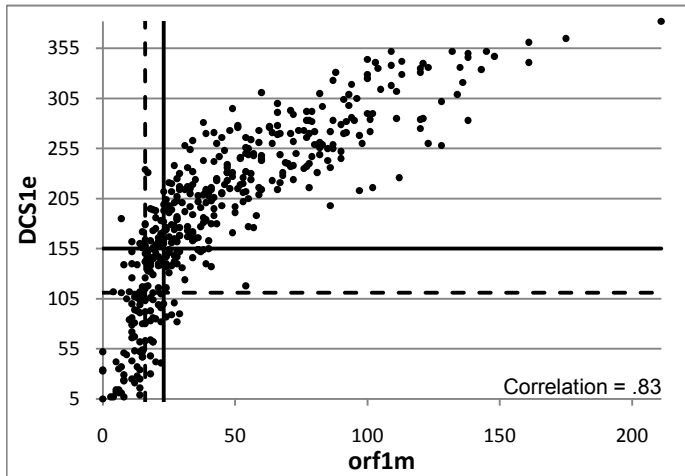
wwr1e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	10	16	113	139	72%
Below Benchmark	7	6	14	27	14%
Well Below Benchmark	14	5	9	28	14%
Marginal Total	31	27	136	194	
Marginal Percent	16%	14%	70%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	32%	59%	83%		

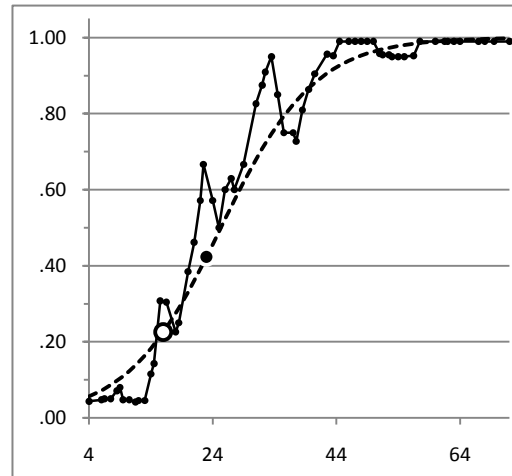
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	113	129	127	149
False Negative	23	34	9	14
True Positive	32	21	19	14
False Positive	26	10	39	17
Sensitivity	.58	.38	.68	.50
specificity	.81	.93	.77	.90
Negative Predictive Power	.83	.79	.93	.91
Positive Predictive Power	.55	.68	.33	.45
Accurate Classification	.75	.77	.75	.84
Kappa	.39	.36	.31	.38

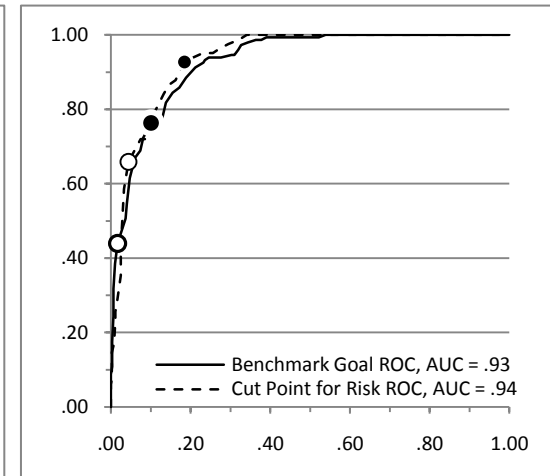
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf1m	23	16	DORF Words Correct, Grade 1, Middle of Year
Outcome	Criterion	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf1m Screening Decision:

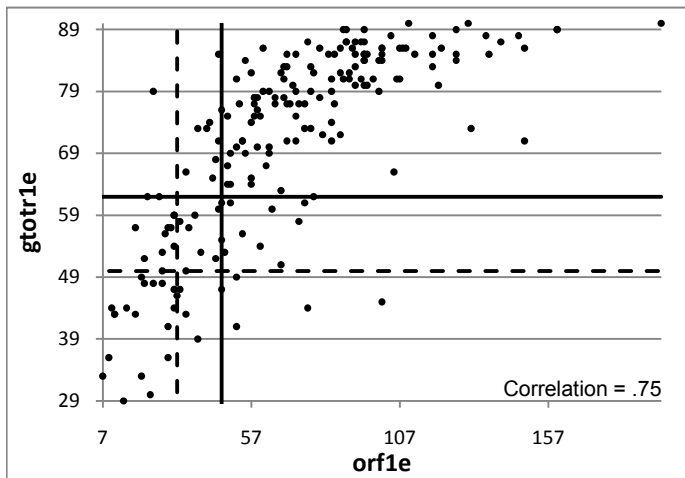
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1e Outcome:					
At or Above Benchmark	5	25	267	297	67%
Below Benchmark	11	26	29	66	15%
Well Below Benchmark	54	22	6	82	18%
Marginal Total	70	73	302	445	
Marginal Percent	16%	16%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	34%	88%		

At or Above Benchmark outcome

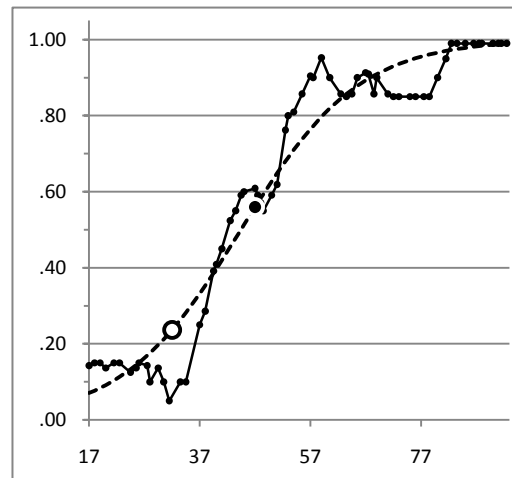
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	267	292	296	347
False Negative	35	83	6	28
True Positive	113	65	76	54
False Positive	30	5	67	16
Sensitivity	.76	.44	.93	.66
specificity	.90	.98	.82	.96
Negative Predictive Power	.88	.78	.98	.93
Positive Predictive Power	.79	.93	.53	.77
Accurate Classification	.85	.80	.84	.90
Kappa	.67	.49	.58	.65

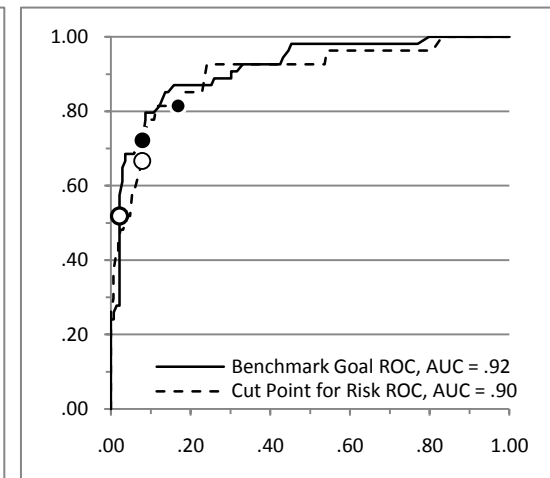
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf1e	47	32	DORF Words Correct, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf1e Screening Decision:

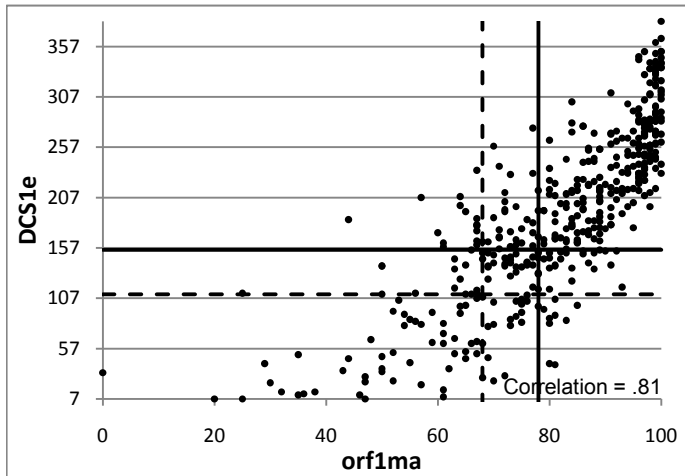
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	3	8	128	139	72%
Below Benchmark	10	7	10	27	14%
Well Below Benchmark	18	4	5	27	14%
Marginal Total	31	19	143	193	
Marginal Percent	16%	10%	74%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	42%	90%		

At or Above Benchmark outcome

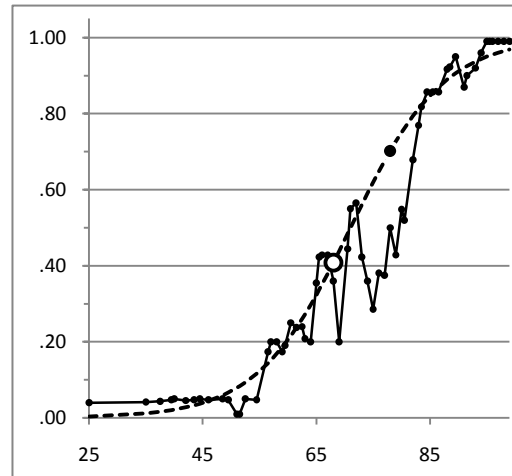
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	128	136	138	153
False Negative	15	26	5	9
True Positive	39	28	22	18
False Positive	11	3	28	13
Sensitivity	.72	.52	.81	.67
specificity	.92	.98	.83	.92
Negative Predictive Power	.90	.84	.97	.94
Positive Predictive Power	.78	.90	.44	.58
Accurate Classification	.87	.85	.83	.89
Kappa	.66	.57	.48	.55

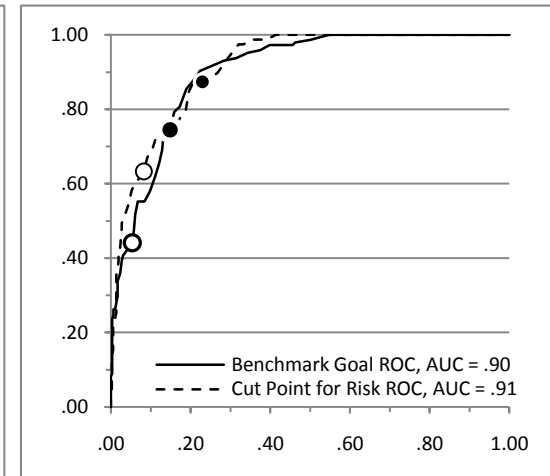
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf1ma	78	68	DORF Accuracy, Grade 1, Middle of Year
Outcome	Criterion	DCS1e	155	111	DIBELS Composite Score, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf1ma Screening Decision:

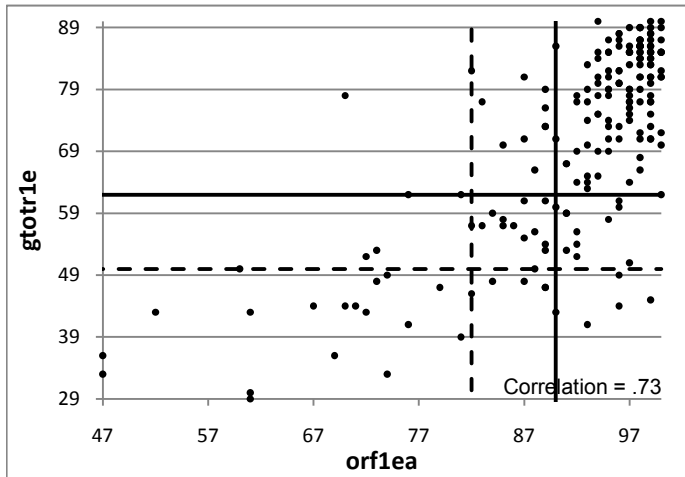
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS1e Outcome:					
At or Above Benchmark	16	28	252	296	67%
Below Benchmark	14	25	27	66	15%
Well Below Benchmark	50	19	10	79	18%
Marginal Total	80	72	289	441	
Marginal Percent	18%	16%	66%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	39%	87%		

At or Above Benchmark outcome

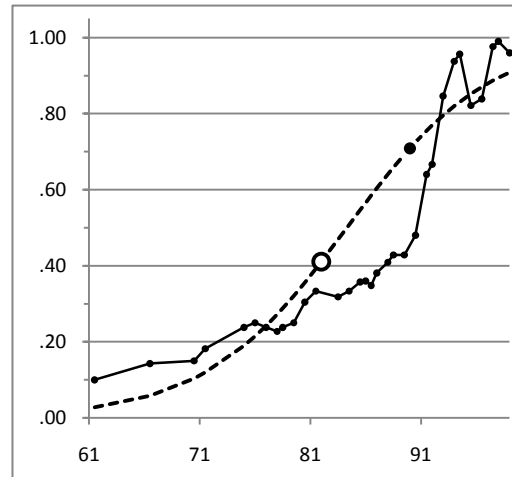
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	252	280	279	332
False Negative	37	81	10	29
True Positive	108	64	69	50
False Positive	44	16	83	30
Sensitivity	.74	.44	.87	.63
specificity	.85	.95	.77	.92
Negative Predictive Power	.87	.78	.97	.92
Positive Predictive Power	.71	.80	.45	.63
Accurate Classification	.82	.78	.79	.87
Kappa	.59	.44	.47	.55

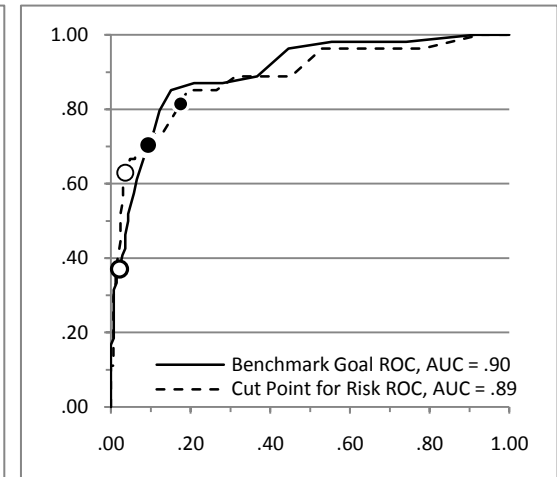
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf1ea	90	82	DORF Accuracy, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf1ea Screening Decision:

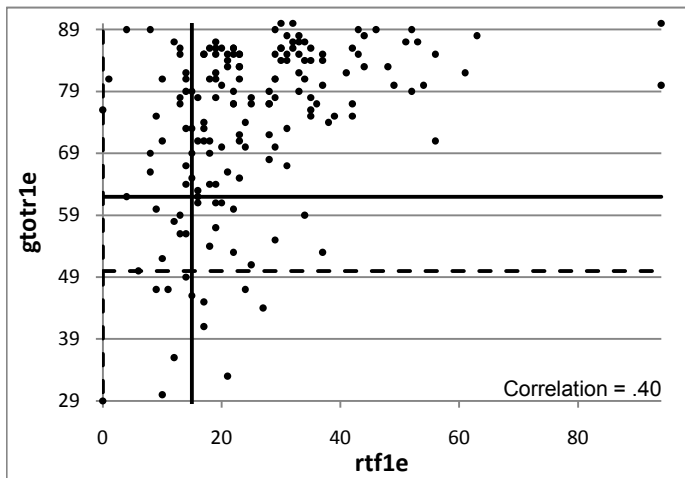
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	3	10	126	139	72%
Below Benchmark	3	13	11	27	14%
Well Below Benchmark	17	5	5	27	14%
Marginal Total	23	28	142	193	
Marginal Percent	12%	15%	74%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	13%	36%	89%		

At or Above Benchmark outcome

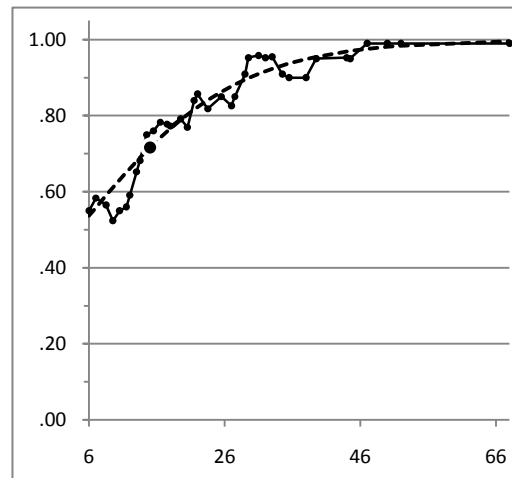
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	126	136	137	160
False Negative	16	34	5	10
True Positive	38	20	22	17
False Positive	13	3	29	6
Sensitivity	.70	.37	.81	.63
specificity	.91	.98	.83	.96
Negative Predictive Power	.89	.80	.96	.94
Positive Predictive Power	.75	.87	.43	.74
Accurate Classification	.85	.81	.82	.92
Kappa	.62	.42	.47	.63

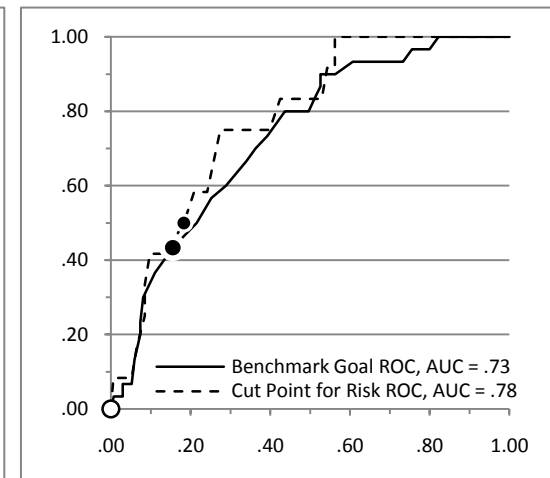
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf1e	15	0	DORF Retell, Grade 1, End of Year
Outcome	Criterion	gtotr1e	62	50	GRADE Total Test, Grade 1, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

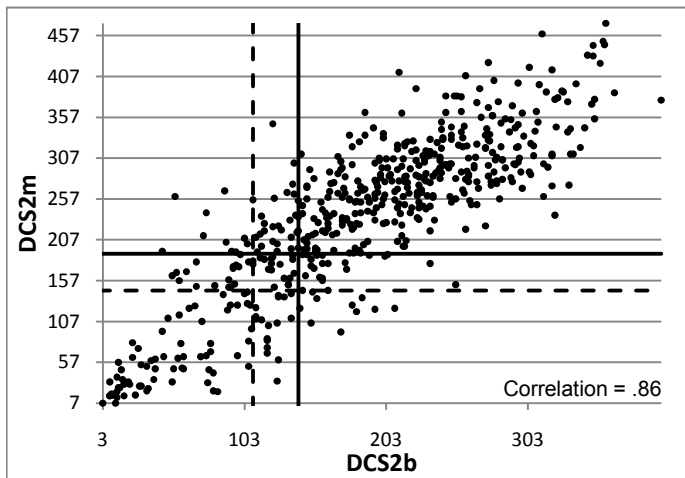
rtf1e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr1e Outcome:					
At or Above Benchmark	0	21	114	135	82%
Below Benchmark	0	7	11	18	11%
Well Below Benchmark	0	6	6	12	7%
Marginal Total	0	34	131	165	
Marginal Percent	0%	21%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	62%	87%		

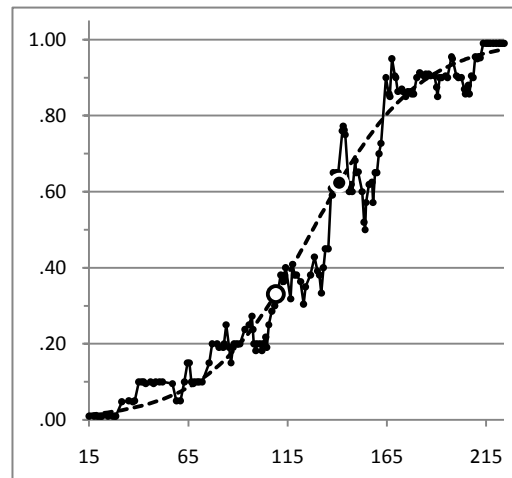
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	114	135	125	153
False Negative	17	30	6	12
True Positive	13	0	6	0
False Positive	21	0	28	0
Sensitivity	.43	.00	.50	.00
specificity	.84	1.00	.82	1.00
Negative Predictive Power	.87	.82	.95	.93
Positive Predictive Power	.38	--	.18	--
Accurate Classification	.77	.82	.79	.93
Kappa	.26	.00	.17	.00

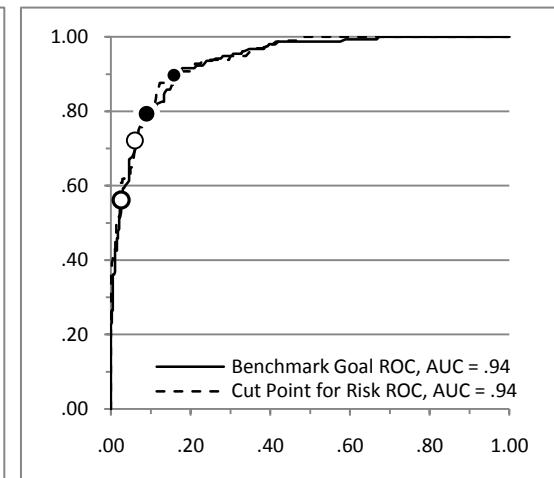
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS2b	141	109	DIBELS Composite Score, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

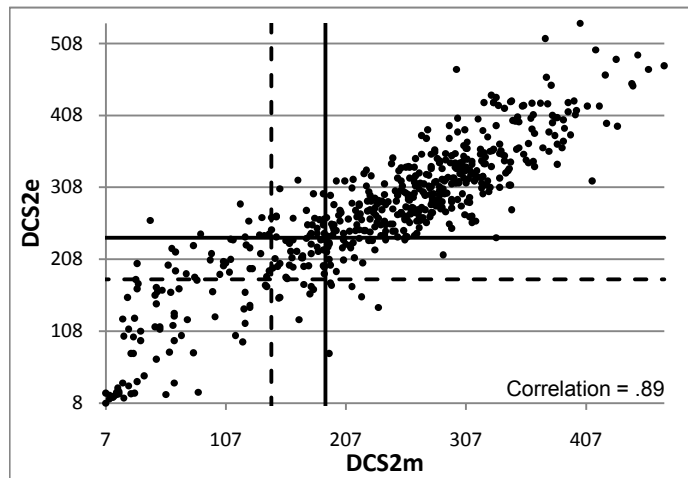
DCS2b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2m Outcome:					
At or Above Benchmark	10	25	356	391	72%
Below Benchmark	17	19	22	58	11%
Well Below Benchmark	70	17	10	97	18%
Marginal Total	97	61	388	546	
Marginal Percent	18%	11%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	41%	92%		

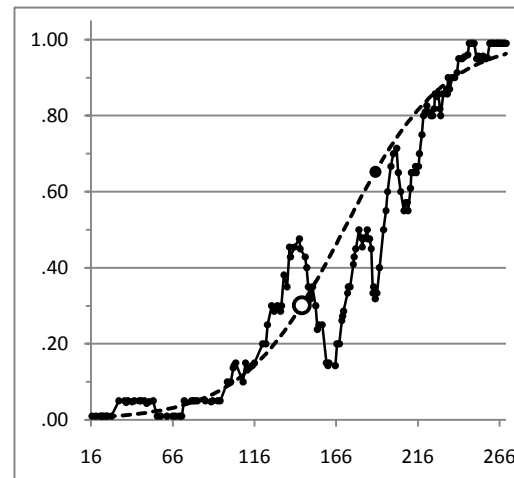
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	356	381	378	422
False Negative	32	68	10	27
True Positive	123	87	87	70
False Positive	35	10	71	27
Sensitivity	.79	.56	.90	.72
specificity	.91	.97	.84	.94
Negative Predictive Power	.92	.85	.97	.94
Positive Predictive Power	.78	.90	.55	.72
Accurate Classification	.88	.86	.85	.90
Kappa	.70	.60	.59	.66

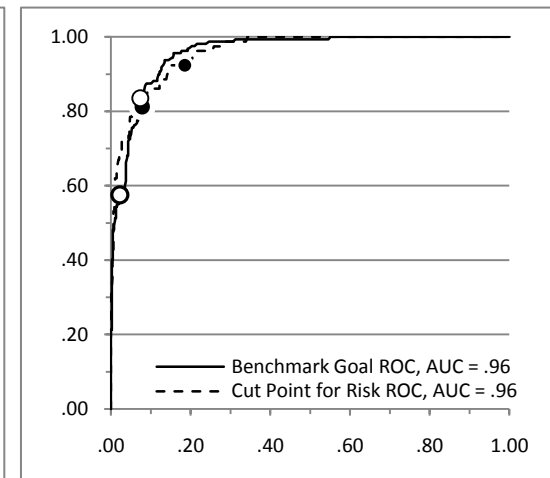
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year
Outcome	Criterion	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

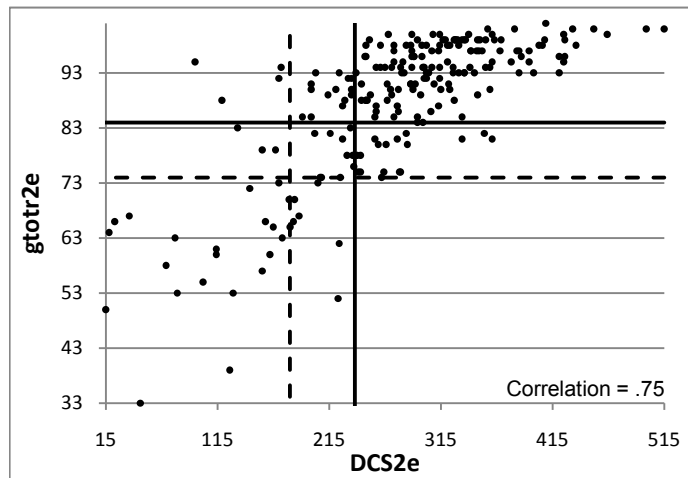
DCS2m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2e Outcome:					
At or Above Benchmark	9	22	363	394	71%
Below Benchmark	26	31	24	81	15%
Well Below Benchmark	66	7	6	79	14%
Marginal Total	101	60	393	554	
Marginal Percent	18%	11%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	9%	37%	92%		

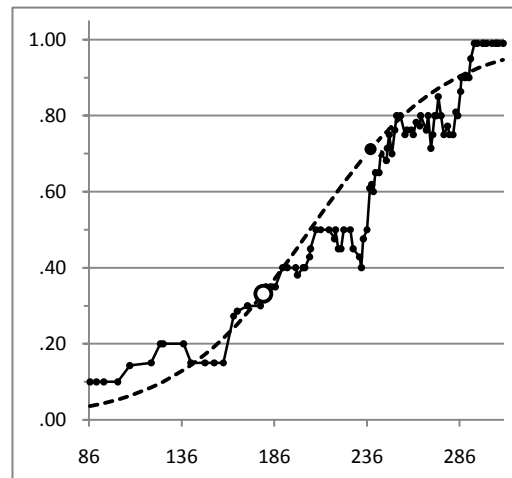
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	363	385	387	440
False Negative	30	68	6	13
True Positive	130	92	73	66
False Positive	31	9	88	35
Sensitivity	.81	.58	.92	.84
specificity	.92	.98	.81	.93
Negative Predictive Power	.92	.85	.98	.97
Positive Predictive Power	.81	.91	.45	.65
Accurate Classification	.89	.86	.83	.91
Kappa	.73	.62	.52	.68

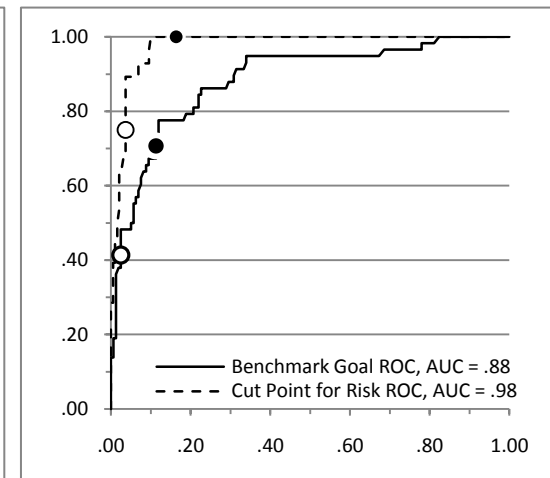
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year
Outcome	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

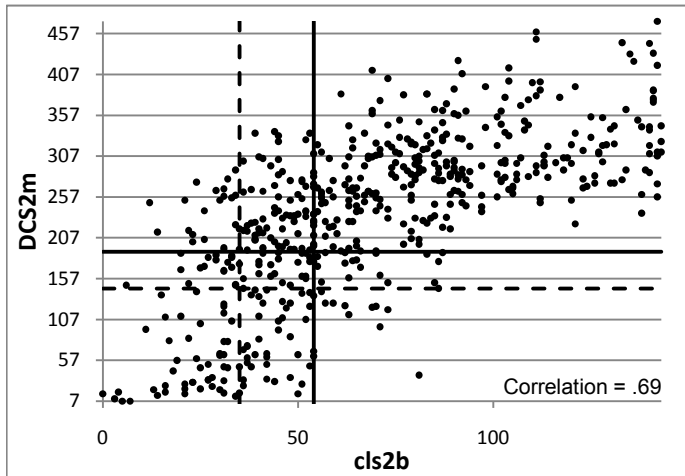
DCS2e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr2e Outcome:					
At or Above Benchmark	4	14	141	159	73%
Below Benchmark	3	10	17	30	14%
Well Below Benchmark	21	7	0	28	13%
Marginal Total	28	31	158	217	
Marginal Percent	13%	14%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	45%	89%		

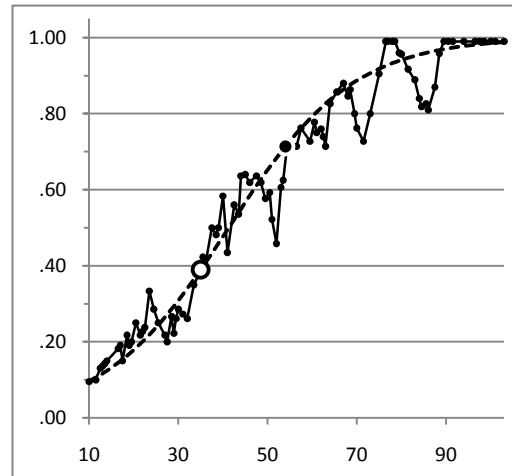
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	141	155	158	182
False Negative	17	34	0	7
True Positive	41	24	28	21
False Positive	18	4	31	7
Sensitivity	.71	.41	1.00	.75
specificity	.89	.97	.84	.96
Negative Predictive Power	.89	.82	1.00	.96
Positive Predictive Power	.69	.86	.47	.75
Accurate Classification	.84	.82	.86	.94
Kappa	.59	.47	.57	.71

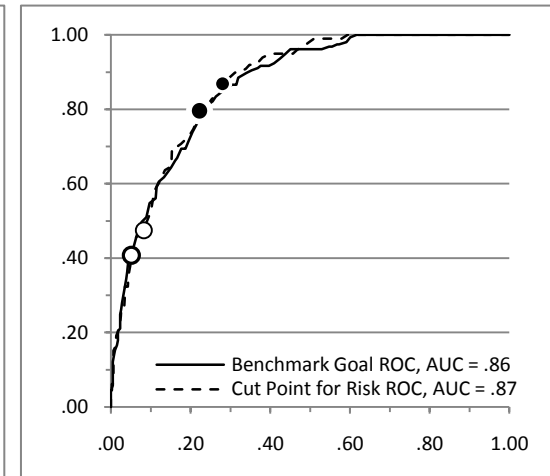
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	cls2b	54	35	NWF Correct Letter Sounds, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

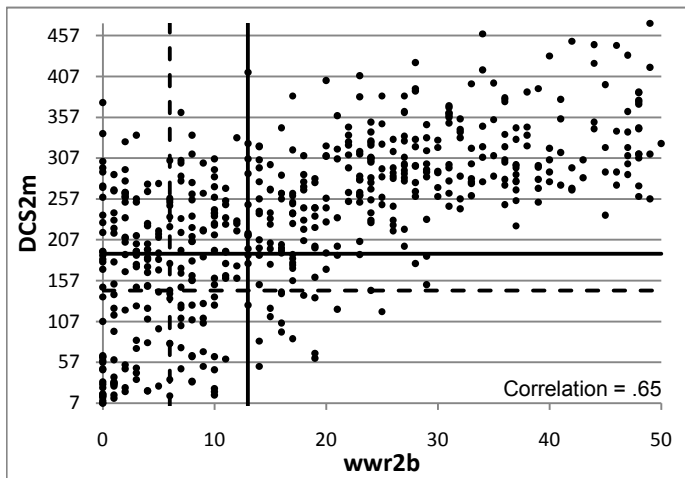
cls2b Screening Decision:

DCS2m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	20	67	304	391	71%
Below Benchmark	17	22	19	58	11%
Well Below Benchmark	47	39	13	99	18%
Marginal Total	84	128	336	548	
Marginal Percent	15%	23%	61%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	24%	52%	90%		

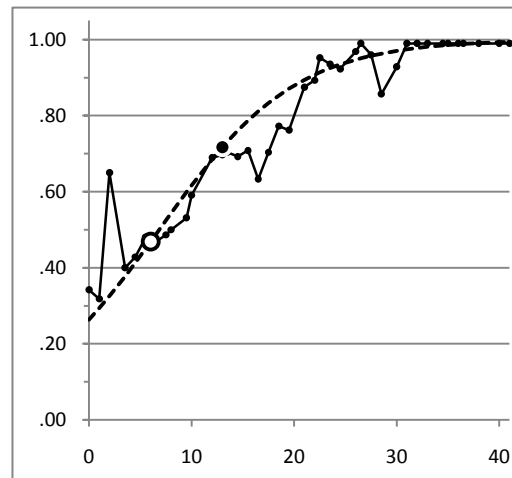
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	304	371	323	412
False Negative	32	93	13	52
True Positive	125	64	86	47
False Positive	87	20	126	37
Sensitivity	.80	.41	.87	.47
specificity	.78	.95	.72	.92
Negative Predictive Power	.90	.80	.96	.89
Positive Predictive Power	.59	.76	.41	.56
Accurate Classification	.78	.79	.75	.84
Kappa	.52	.41	.41	.42

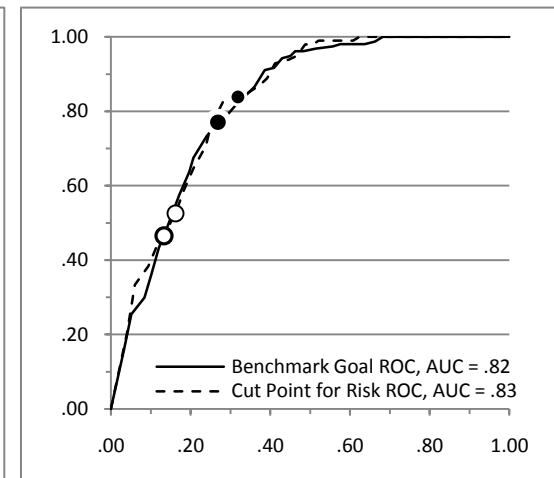
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	wwr2b	13	6	NWF Whole Words Read, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

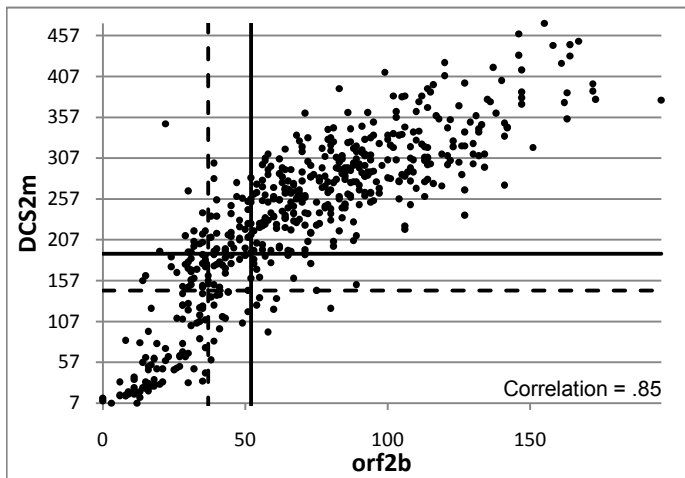
wwr2b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2m Outcome:					
At or Above Benchmark	52	53	286	391	71%
Below Benchmark	21	17	20	58	11%
Well Below Benchmark	52	31	16	99	18%
Marginal Total	125	101	322	548	
Marginal Percent	23%	18%	59%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	42%	52%	89%		

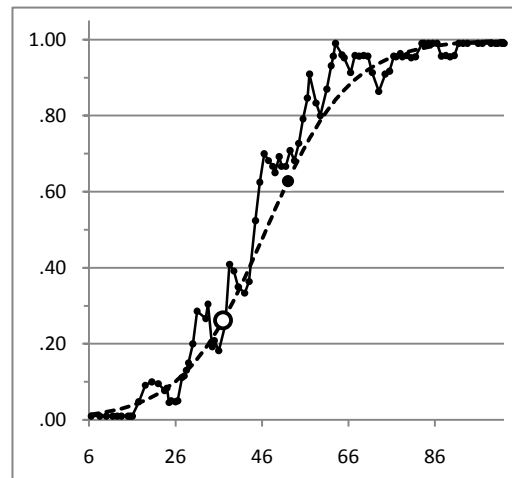
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	286	339	306	376
False Negative	36	84	16	47
True Positive	121	73	83	52
False Positive	105	52	143	73
Sensitivity	.77	.46	.84	.53
specificity	.73	.87	.68	.84
Negative Predictive Power	.89	.80	.95	.89
Positive Predictive Power	.54	.58	.37	.42
Accurate Classification	.74	.75	.71	.78
Kappa	.44	.35	.35	.33

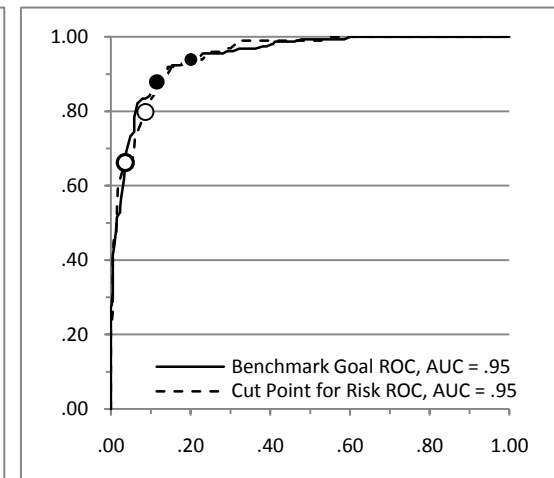
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2b	52	37	DORF Words Correct, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

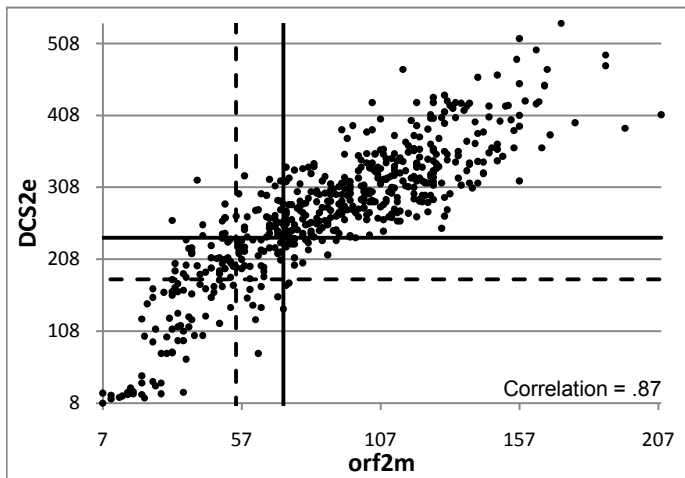
orf2b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2m Outcome:					
At or Above Benchmark	14	31	346	391	71%
Below Benchmark	25	20	13	58	11%
Well Below Benchmark	79	14	6	99	18%
Marginal Total	118	65	365	548	
Marginal Percent	22%	12%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	48%	95%		

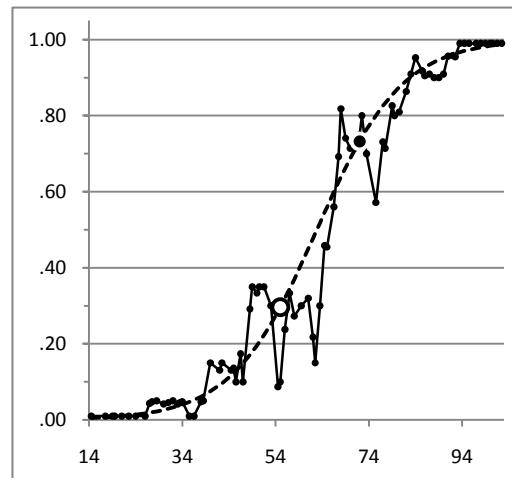
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	346	377	359	410
False Negative	19	53	6	20
True Positive	138	104	93	79
False Positive	45	14	90	39
Sensitivity	.88	.66	.94	.80
specificity	.88	.96	.80	.91
Negative Predictive Power	.95	.88	.98	.95
Positive Predictive Power	.75	.88	.51	.67
Accurate Classification	.88	.88	.82	.89
Kappa	.73	.68	.56	.66

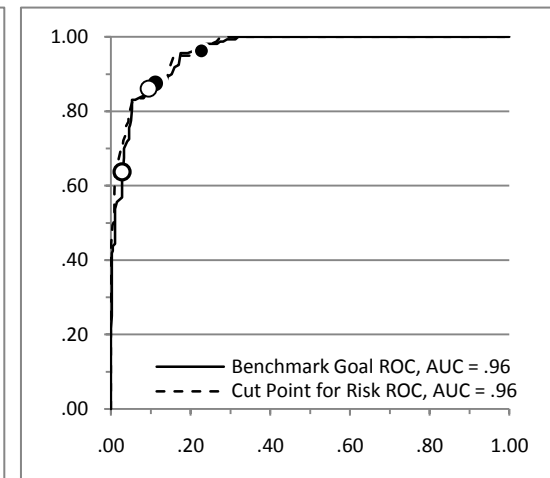
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2m	72	55	DORF Words Correct, Grade 2, Middle of Year
Outcome	Criterion	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

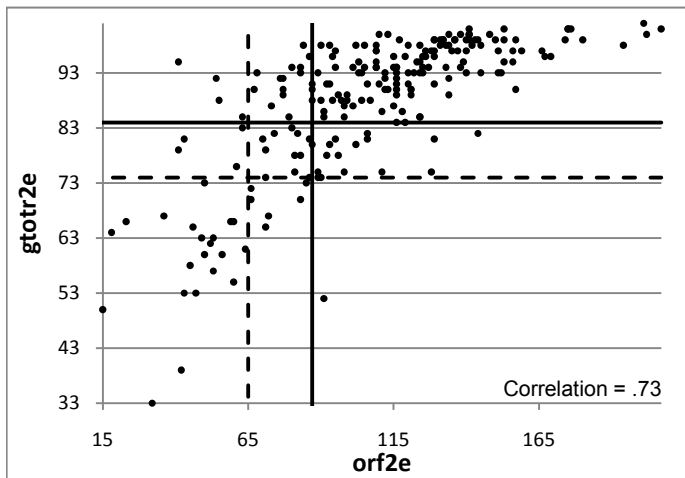
orf2m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2e Outcome:					
At or Above Benchmark	11	33	350	394	71%
Below Benchmark	34	30	17	81	15%
Well Below Benchmark	68	8	3	79	14%
Marginal Total	113	71	370	554	
Marginal Percent	20%	13%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	46%	95%		

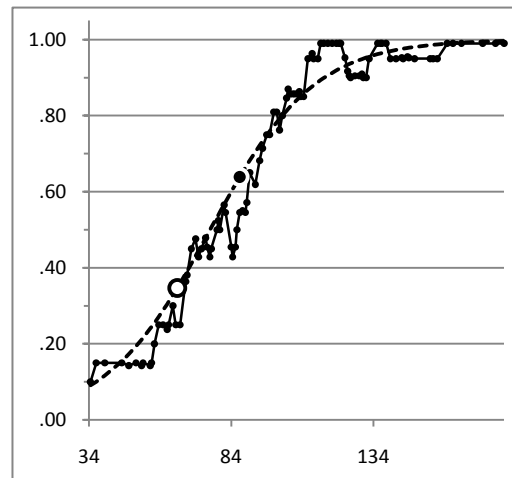
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	350	383	367	430
False Negative	20	58	3	11
True Positive	140	102	76	68
False Positive	44	11	108	45
Sensitivity	.88	.64	.96	.86
specificity	.89	.97	.77	.91
Negative Predictive Power	.95	.87	.99	.98
Positive Predictive Power	.76	.90	.41	.60
Accurate Classification	.88	.88	.80	.90
Kappa	.73	.67	.47	.65

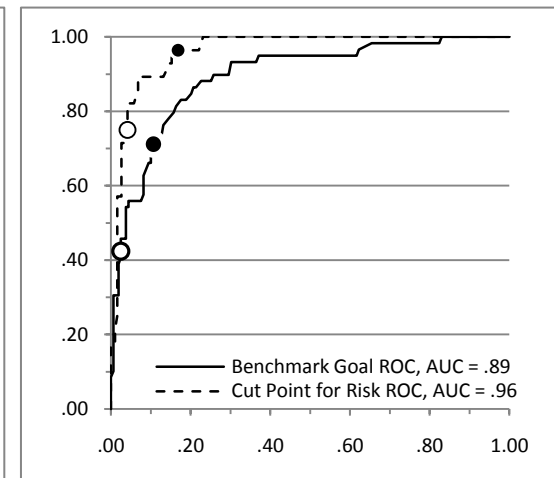
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2e	87	65	DORF Words Correct, Grade 2, End of Year
Outcome	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

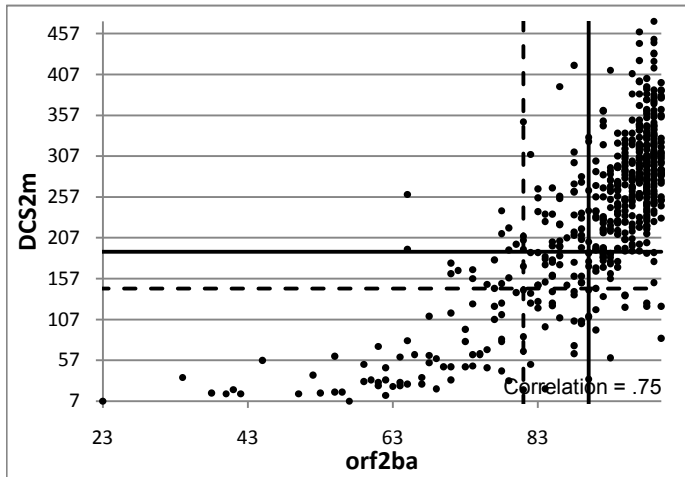
orf2e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr2e Outcome:					
At or Above Benchmark	4	13	142	159	73%
Below Benchmark	4	11	16	31	14%
Well Below Benchmark	21	6	1	28	13%
Marginal Total	29	30	159	218	
Marginal Percent	13%	14%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	43%	89%		

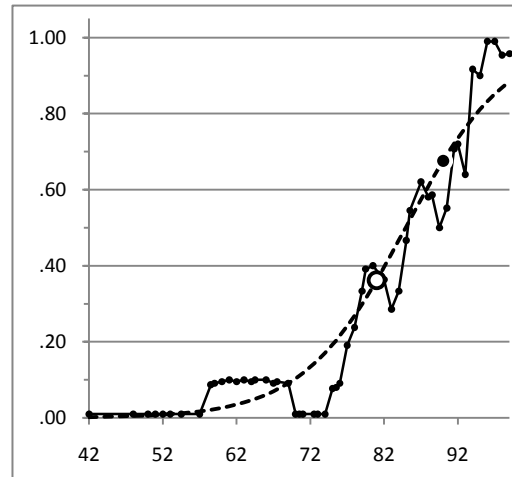
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	142	155	158	182
False Negative	17	34	1	7
True Positive	42	25	27	21
False Positive	17	4	32	8
Sensitivity	.71	.42	.96	.75
specificity	.89	.97	.83	.96
Negative Predictive Power	.89	.82	.99	.96
Positive Predictive Power	.71	.86	.46	.72
Accurate Classification	.84	.83	.85	.93
Kappa	.60	.47	.54	.70

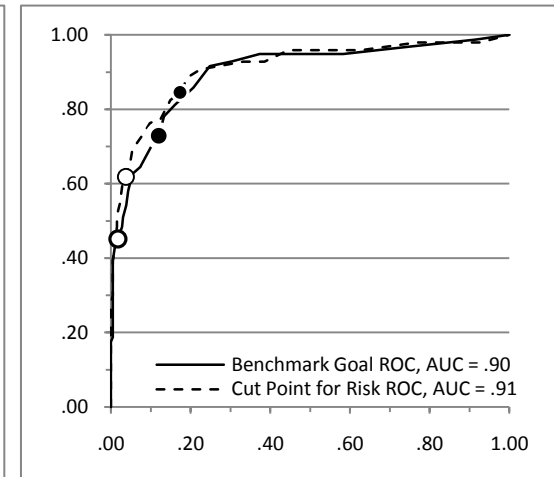
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2ba	90	81	DORF Accuracy, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

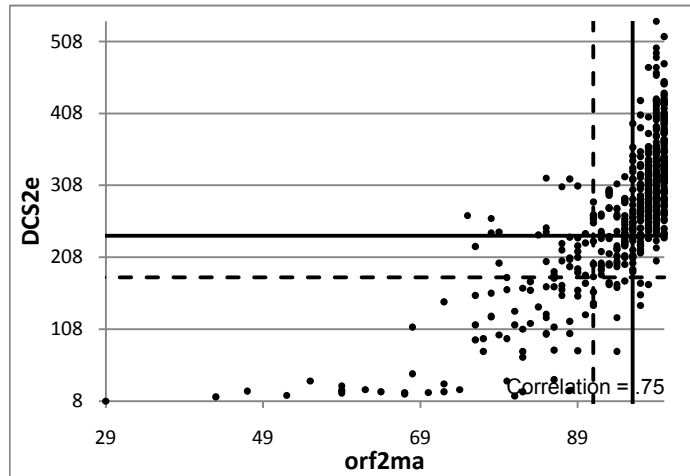
orf2ba Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2m Outcome:					
At or Above Benchmark	7	40	344	391	72%
Below Benchmark	10	21	27	58	11%
Well Below Benchmark	60	22	15	97	18%
Marginal Total	77	83	386	546	
Marginal Percent	14%	15%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	9%	48%	89%		

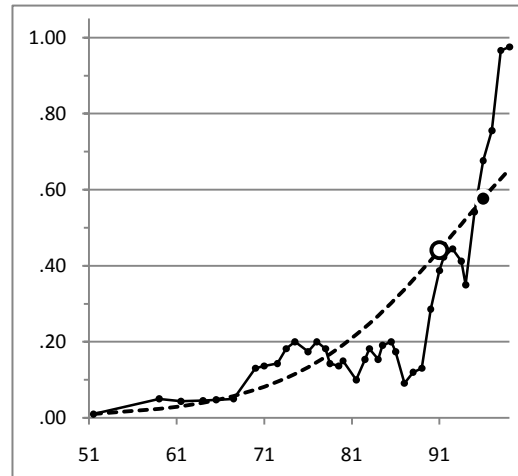
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	344	384	371	432
False Negative	42	85	15	37
True Positive	113	70	82	60
False Positive	47	7	78	17
Sensitivity	.73	.45	.85	.62
specificity	.88	.98	.83	.96
Negative Predictive Power	.89	.82	.96	.92
Positive Predictive Power	.71	.91	.51	.78
Accurate Classification	.84	.83	.83	.90
Kappa	.60	.51	.54	.63

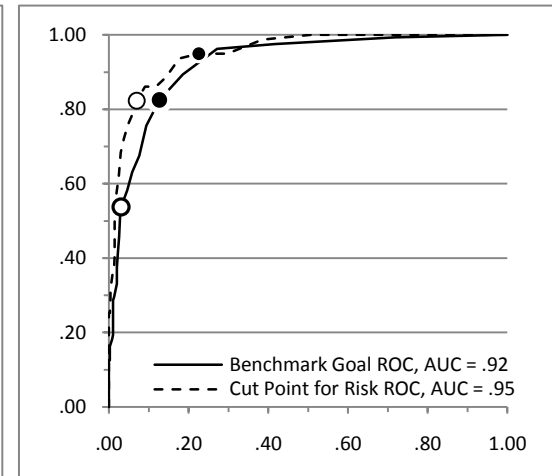
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2ma	96	91	DORF Accuracy, Grade 2, Middle of Year
Outcome	Criterion	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf2ma Screening Decision:

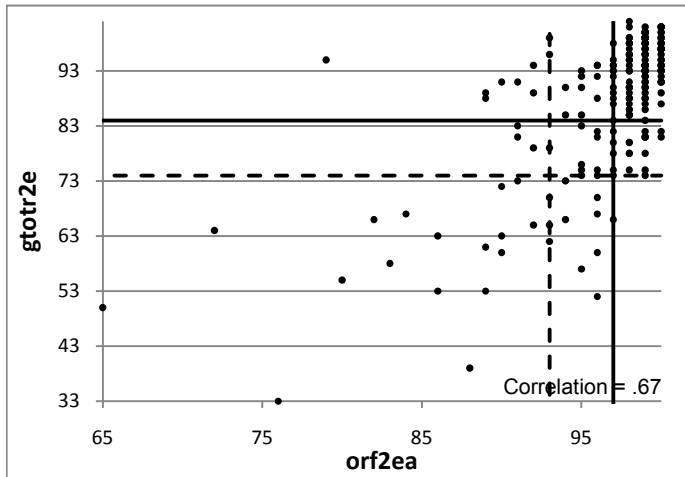
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2e Outcome:					
At or Above Benchmark	12	38	344	394	71%
Below Benchmark	21	36	24	81	15%
Well Below Benchmark	65	10	4	79	14%
Marginal Total	98	84	372	554	
Marginal Percent	18%	15%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	45%	92%		

At or Above Benchmark outcome

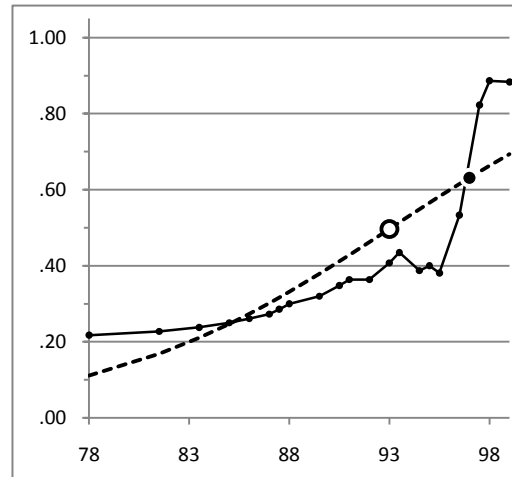
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	344	382	368	442
False Negative	28	74	4	14
True Positive	132	86	75	65
False Positive	50	12	107	33
Sensitivity	.83	.54	.95	.82
specificity	.87	.97	.77	.93
Negative Predictive Power	.92	.84	.99	.97
Positive Predictive Power	.73	.88	.41	.66
Accurate Classification	.86	.84	.80	.92
Kappa	.67	.57	.47	.68

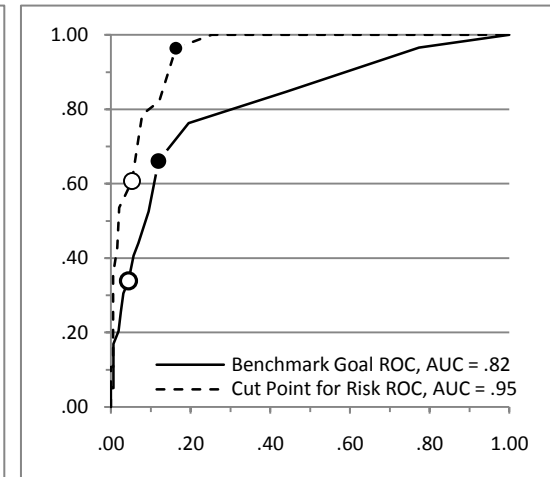
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf2ea	97	93	DORF Accuracy, Grade 2, End of Year
Outcome	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf2ea Screening Decision:

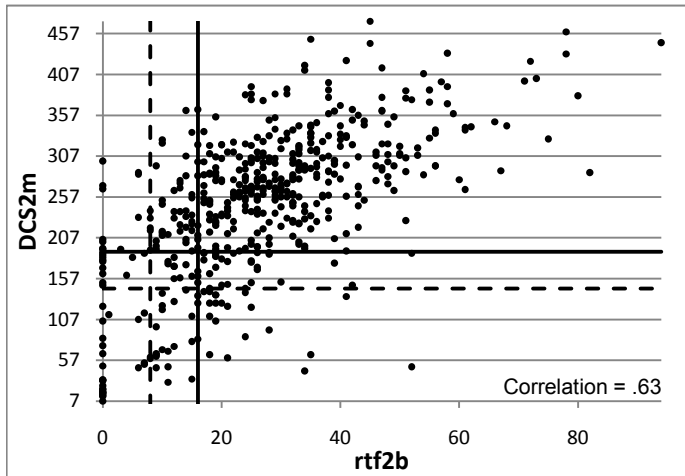
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr2e Outcome:					
At or Above Benchmark	7	12	140	159	73%
Below Benchmark	3	9	19	31	14%
Well Below Benchmark	17	10	1	28	13%
Marginal Total	27	31	160	218	
Marginal Percent	12%	14%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	39%	88%		

At or Above Benchmark outcome

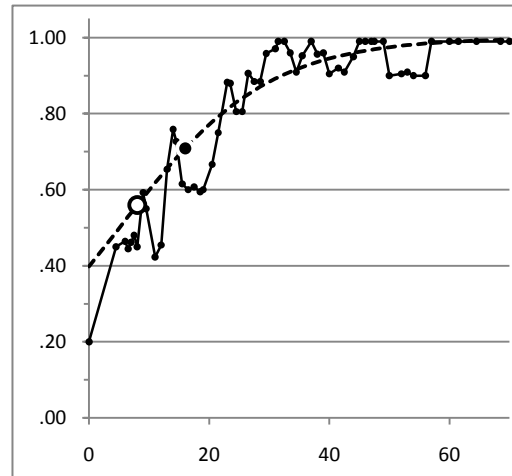
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	140	152	159	180
False Negative	20	39	1	11
True Positive	39	20	27	17
False Positive	19	7	31	10
Sensitivity	.66	.34	.96	.61
specificity	.88	.96	.84	.95
Negative Predictive Power	.88	.80	.99	.94
Positive Predictive Power	.67	.74	.47	.63
Accurate Classification	.82	.79	.85	.90
Kappa	.54	.36	.55	.56

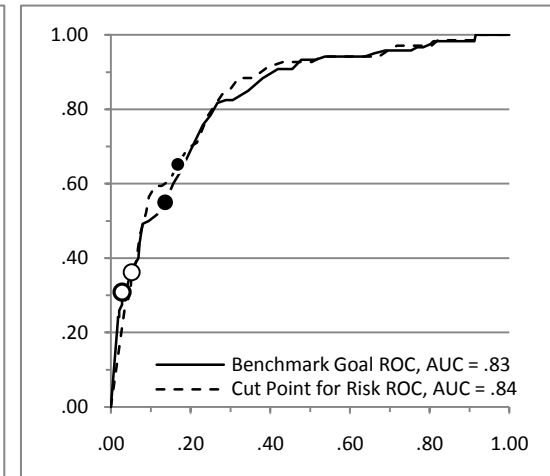
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf2b	16	8	DORF Retell, Grade 2, Beginning of Year
Outcome	Criterion	DCS2m	190	145	DIBELS Composite Score, Grade 2, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

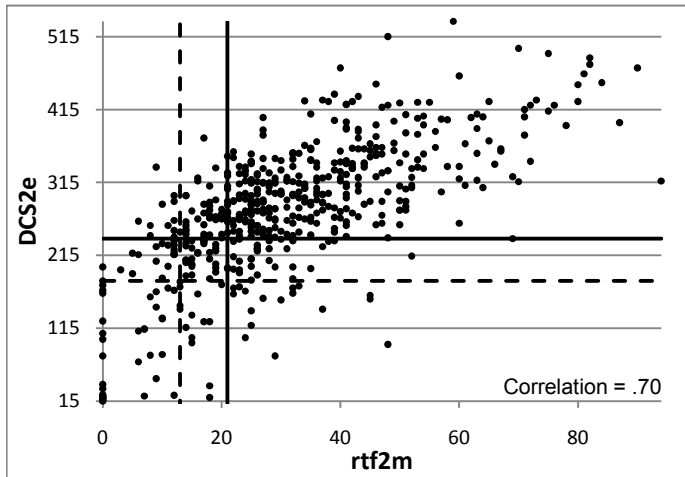
rtf2b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2m Outcome:					
At or Above Benchmark	11	42	336	389	76%
Below Benchmark	12	9	30	51	10%
Well Below Benchmark	25	20	24	69	14%
Marginal Total	48	71	390	509	
Marginal Percent	9%	14%	77%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	23%	59%	86%		

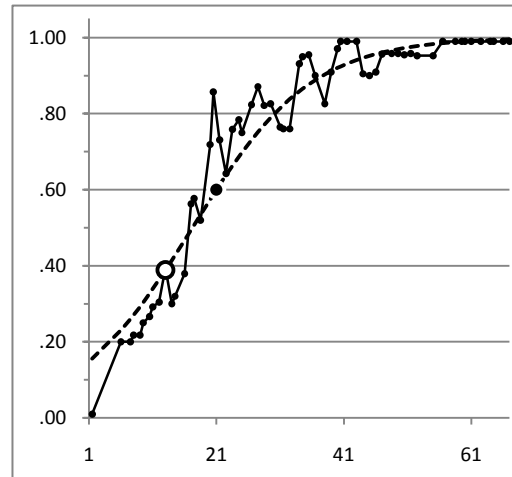
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	336	378	366	417
False Negative	54	83	24	44
True Positive	66	37	45	25
False Positive	53	11	74	23
Sensitivity	.55	.31	.65	.36
specificity	.86	.97	.83	.95
Negative Predictive Power	.86	.82	.94	.90
Positive Predictive Power	.55	.77	.38	.52
Accurate Classification	.79	.82	.81	.87
Kappa	.41	.35	.37	.36

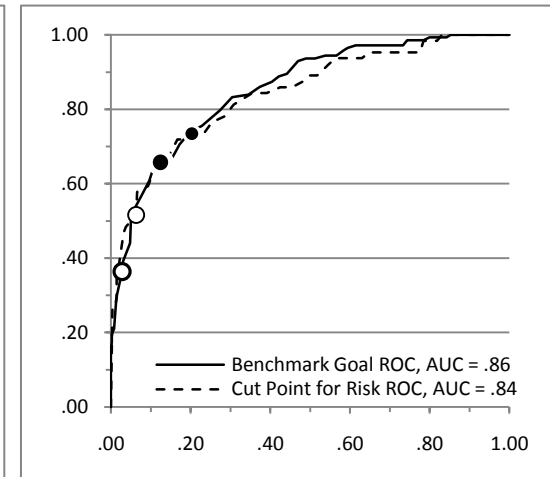
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf2m	21	13	DORF Retell, Grade 2, Middle of Year
Outcome	Criterion	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rtf2m Screening Decision:

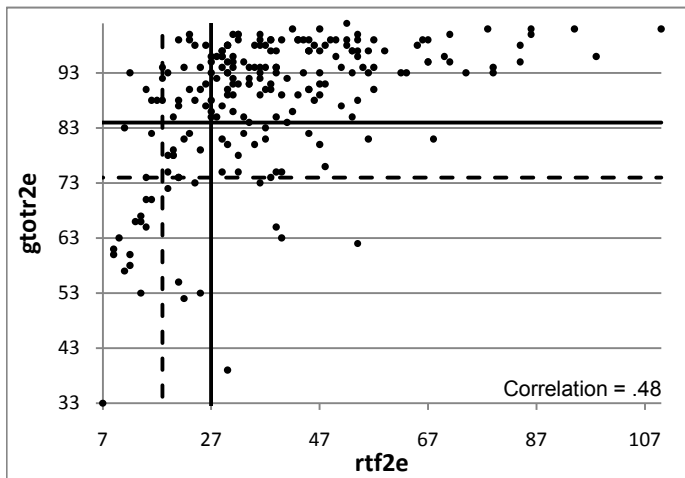
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2e Outcome:					
At or Above Benchmark	11	38	345	394	73%
Below Benchmark	19	28	32	79	15%
Well Below Benchmark	33	14	17	64	12%
Marginal Total	63	80	394	537	
Marginal Percent	12%	15%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	17%	48%	88%		

At or Above Benchmark outcome

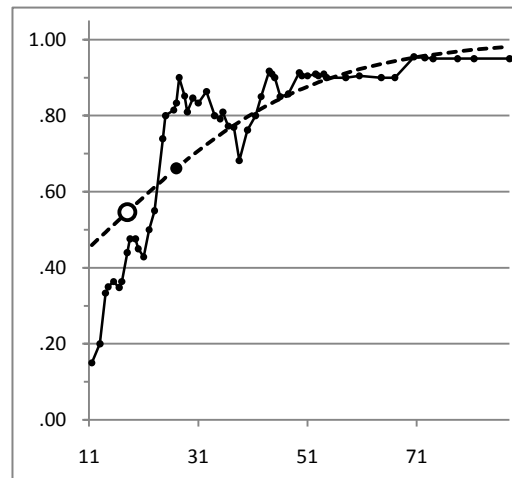
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	345	383	377	443
False Negative	49	91	17	31
True Positive	94	52	47	33
False Positive	49	11	96	30
Sensitivity	.66	.36	.73	.52
specificity	.88	.97	.80	.94
Negative Predictive Power	.88	.81	.96	.93
Positive Predictive Power	.66	.83	.33	.52
Accurate Classification	.82	.81	.79	.89
Kappa	.53	.41	.35	.46

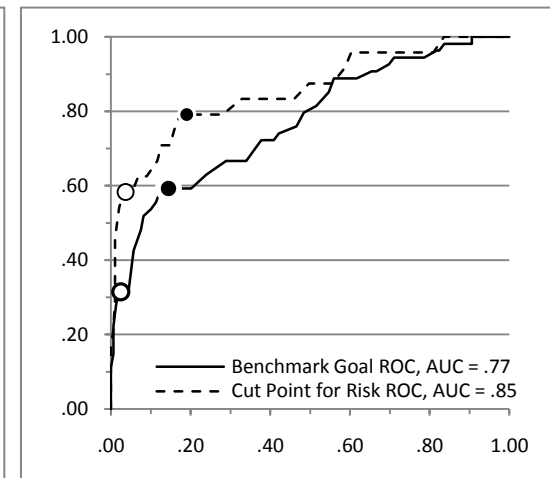
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf2e	27	18	DORF Retell, Grade 2, End of Year
Outcome	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

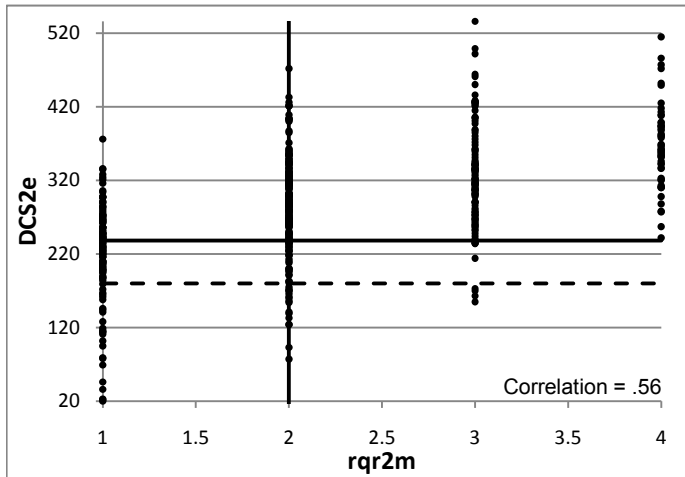
rtf2e Screening Decision:

gtotr2e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	19	136	159	75%
Below Benchmark	3	10	17	30	14%
Well Below Benchmark	14	5	5	24	11%
Marginal Total	21	34	158	213	
Marginal Percent	10%	16%	74%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	19%	56%	86%		

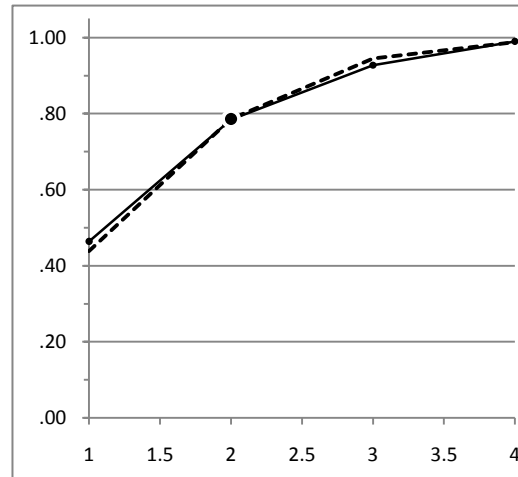
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	136	155	153	182
False Negative	22	37	5	10
True Positive	32	17	19	14
False Positive	23	4	36	7
Sensitivity	.59	.31	.79	.58
specificity	.86	.97	.81	.96
Negative Predictive Power	.86	.81	.97	.95
Positive Predictive Power	.58	.81	.35	.67
Accurate Classification	.79	.81	.81	.92
Kappa	.45	.36	.38	.58

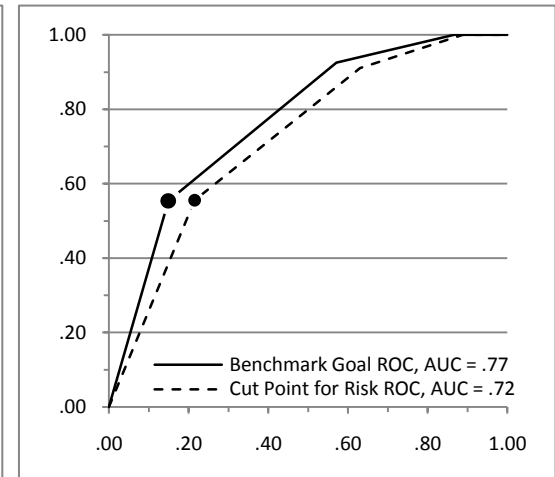
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr2m	2	0	DORF Retell Quality of Response, Grade 2, Middle of Year
Outcome	Criterion	DCS2e	238	180	DIBELS Composite Score, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

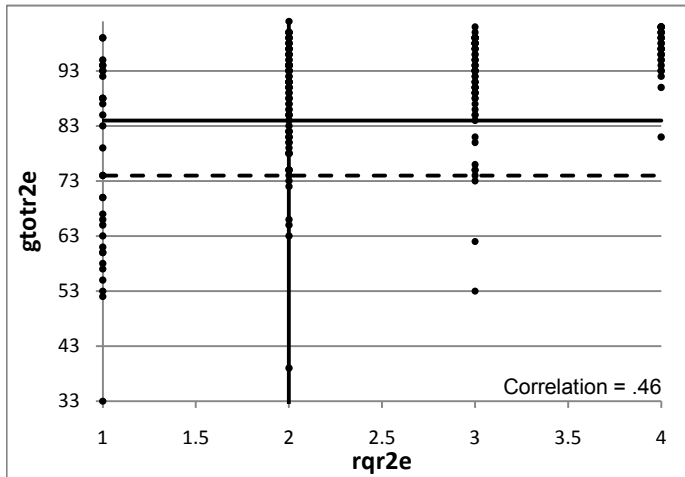
rqr2m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS2e Outcome:					
At or Above Benchmark	0	58	331	389	76%
Below Benchmark	0	42	34	76	15%
Well Below Benchmark	0	25	20	45	9%
Marginal Total	0	125	385	510	
Marginal Percent	0%	25%	75%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	46%	86%		

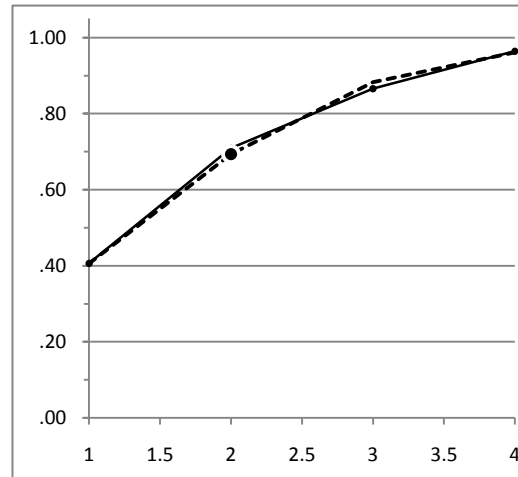
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	331	389	365	465
False Negative	54	121	20	45
True Positive	67	0	25	0
False Positive	58	0	100	0
Sensitivity	.55	.00	.56	.00
specificity	.85	1.00	.78	1.00
Negative Predictive Power	.86	.76	.95	.91
Positive Predictive Power	.54	--	.20	--
Accurate Classification	.78	.76	.76	.91
Kappa	.40	.00	.19	.00

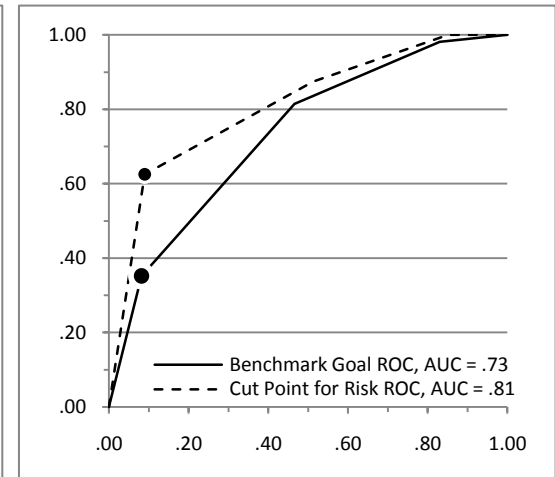
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr2e	2	0	DORF Retell Quality of Response, Grade 2, End of Year
Outcome	Criterion	gtotr2e	84	74	GRADE Total Test, Grade 2, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

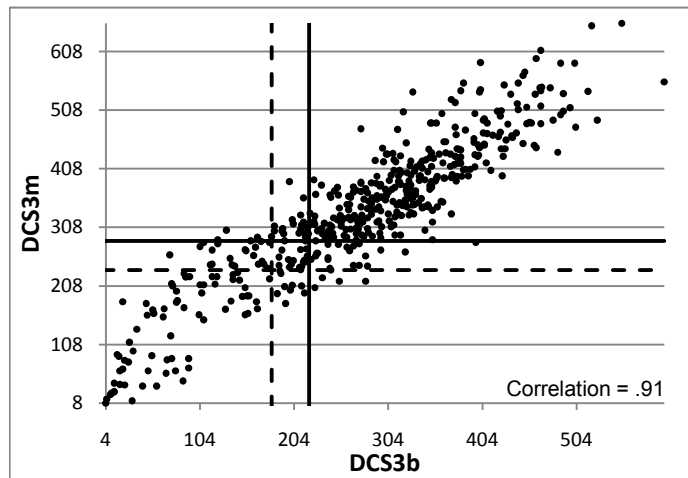
rqr2e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr2e Outcome:					
At or Above Benchmark	0	13	146	159	75%
Below Benchmark	0	4	26	30	14%
Well Below Benchmark	0	15	9	24	11%
Marginal Total	0	32	181	213	
Marginal Percent	0%	15%	85%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	41%	81%		

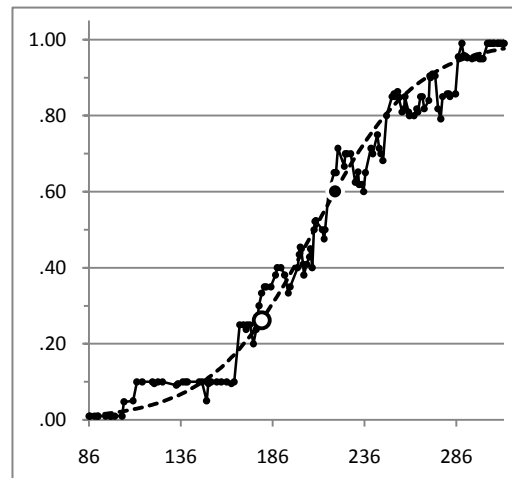
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	146	159	172	189
False Negative	35	54	9	24
True Positive	19	0	15	0
False Positive	13	0	17	0
Sensitivity	.35	.00	.63	.00
specificity	.92	1.00	.91	1.00
Negative Predictive Power	.81	.75	.95	.89
Positive Predictive Power	.59	--	.47	--
Accurate Classification	.77	.75	.88	.89
Kappa	.31	.00	.47	.00

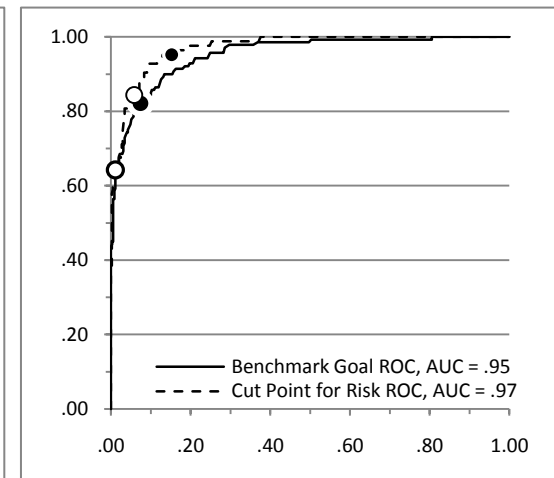
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS3b	220	180	DIBELS Composite Score, Grade 3, Beginning of Year
Outcome	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

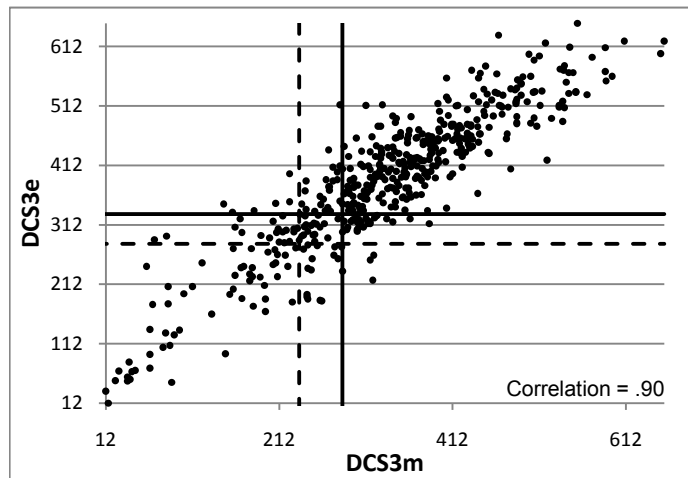
DCS3b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3m Outcome:					
At or Above Benchmark	4	22	324	350	71%
Below Benchmark	20	16	21	57	12%
Well Below Benchmark	70	9	4	83	17%
Marginal Total	94	47	349	490	
Marginal Percent	19%	10%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	4%	47%	93%		

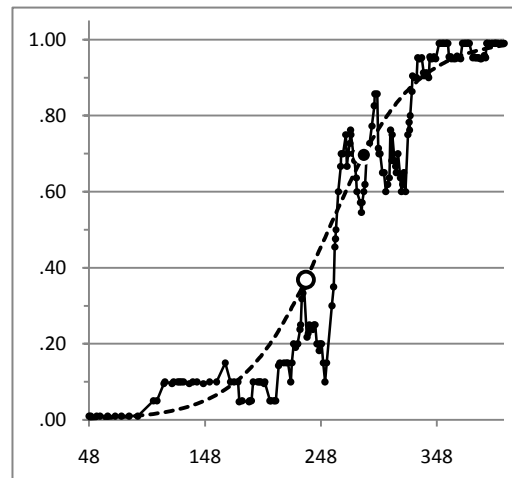
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	324	346	345	383
False Negative	25	50	4	13
True Positive	115	90	79	70
False Positive	26	4	62	24
Sensitivity	.82	.64	.95	.84
Specificity	.93	.99	.85	.94
Negative Predictive Power	.93	.87	.99	.97
Positive Predictive Power	.82	.96	.56	.74
Accurate Classification	.90	.89	.87	.92
Kappa	.75	.70	.63	.75

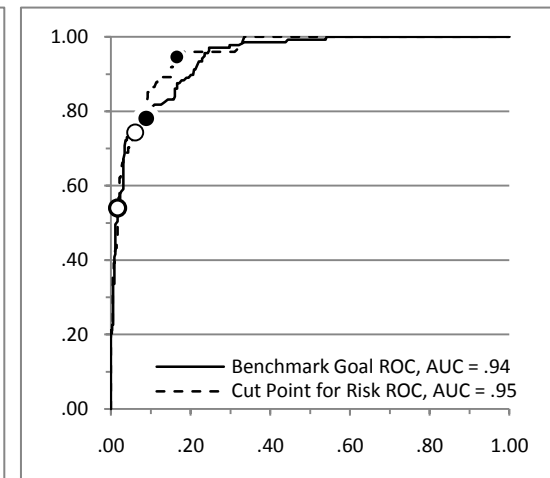
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year
Outcome	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

DCS3m Screening Decision:

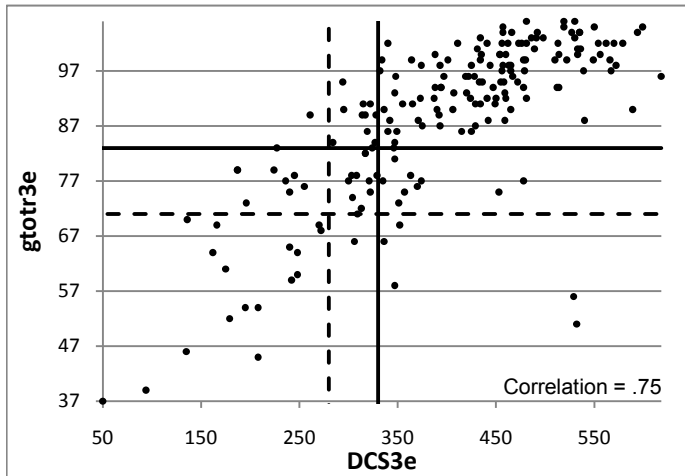
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3e Outcome:					
At or Above Benchmark	6	25	318	349	72%
Below Benchmark	19	18	26	63	13%
Well Below Benchmark	55	15	4	74	15%
Marginal Total	80	58	348	486	
Marginal Percent	16%	12%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	8%	43%	91%		

At or Above Benchmark outcome

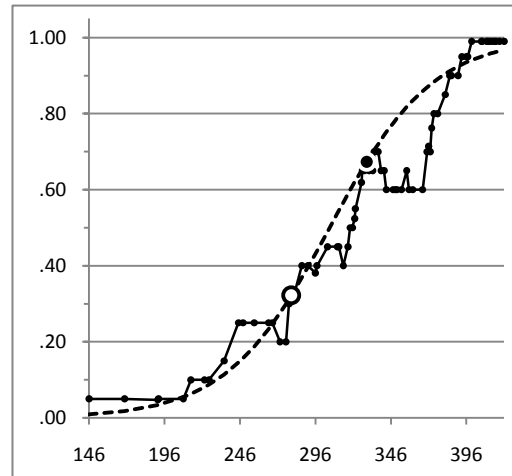
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	318	343	344	387
False Negative	30	63	4	19
True Positive	107	74	70	55
False Positive	31	6	68	25
Sensitivity	.78	.54	.95	.74
Specificity	.91	.98	.83	.94
Negative Predictive Power	.91	.84	.99	.95
Positive Predictive Power	.78	.93	.51	.69
Accurate Classification	.87	.86	.85	.91
Kappa	.69	.60	.58	.66

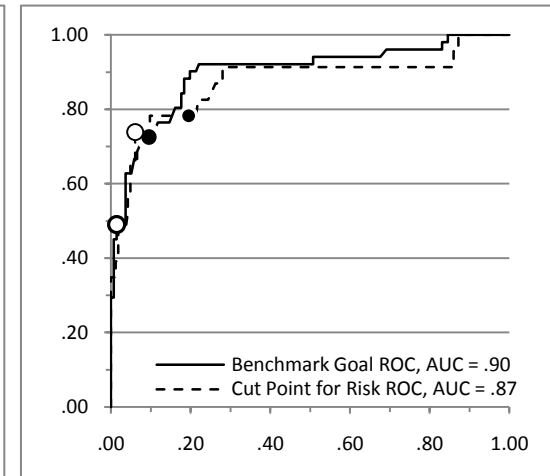
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

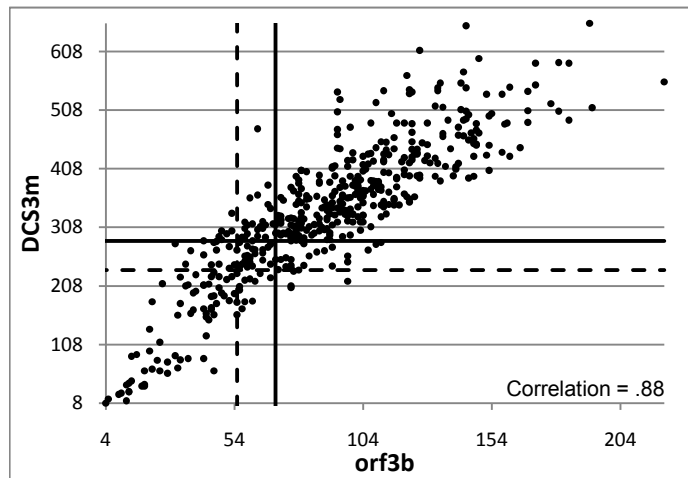
DCS3e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr3e Outcome:					
At or Above Benchmark	2	11	123	136	73%
Below Benchmark	8	11	9	28	15%
Well Below Benchmark	17	1	5	23	12%
Marginal Total	27	23	137	187	
Marginal Percent	14%	12%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	48%	90%		

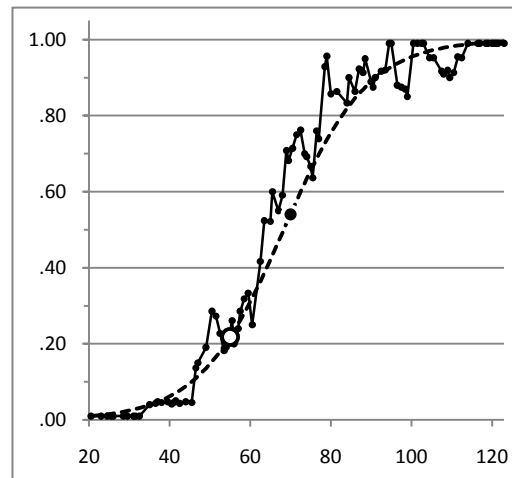
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	123	134	132	154
False Negative	14	26	5	6
True Positive	37	25	18	17
False Positive	13	2	32	10
Sensitivity	.73	.49	.78	.74
Specificity	.90	.99	.80	.94
Negative Predictive Power	.90	.84	.96	.96
Positive Predictive Power	.74	.93	.36	.63
Accurate Classification	.86	.85	.80	.91
Kappa	.63	.56	.39	.63

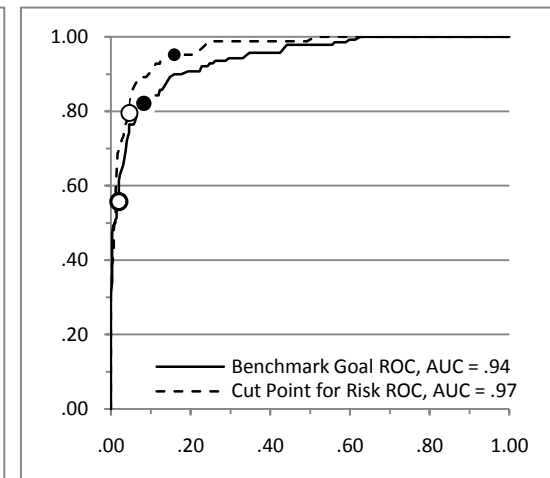
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3b	70	55	DORF Words Correct, Grade 3, Beginning of Year
Outcome	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

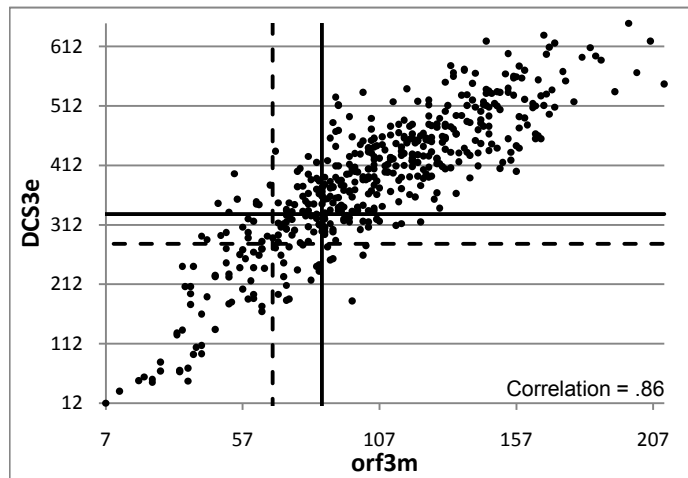
orf3b Screening Decision:

DCS3m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	7	22	322	351	71%
Below Benchmark	12	24	21	57	12%
Well Below Benchmark	66	13	4	83	17%
Marginal Total	85	59	347	491	
Marginal Percent	17%	12%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	8%	37%	93%		

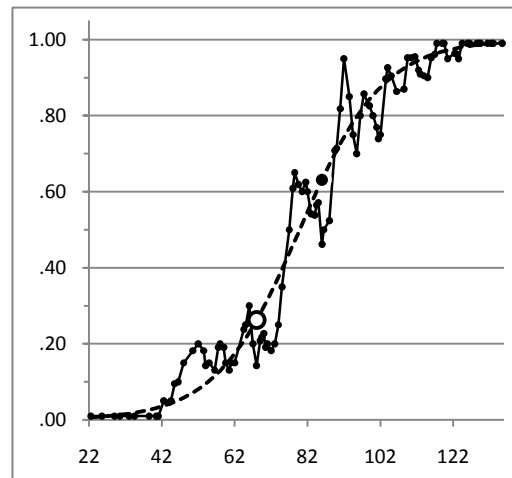
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	322	344	343	389
False Negative	25	62	4	17
True Positive	115	78	79	66
False Positive	29	7	65	19
Sensitivity	.82	.56	.95	.80
Specificity	.92	.98	.84	.95
Negative Predictive Power	.93	.85	.99	.96
Positive Predictive Power	.80	.92	.55	.78
Accurate Classification	.89	.86	.86	.93
Kappa	.73	.61	.61	.74

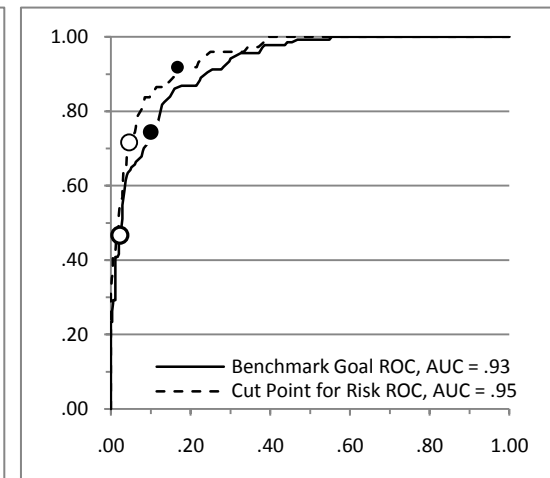
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3m	86	68	DORF Words Correct, Grade 3, Middle of Year
Outcome	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

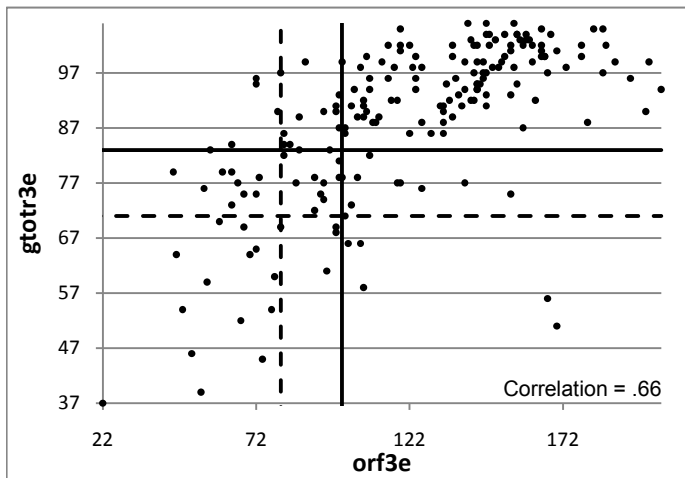
orf3m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3e Outcome:					
At or Above Benchmark	8	27	315	350	72%
Below Benchmark	11	23	29	63	13%
Well Below Benchmark	53	15	6	74	15%
Marginal Total	72	65	350	487	
Marginal Percent	15%	13%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	11%	42%	90%		

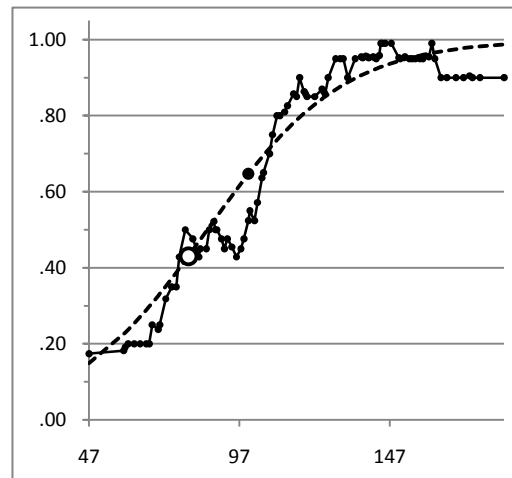
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	315	342	344	394
False Negative	35	73	6	21
True Positive	102	64	68	53
False Positive	35	8	69	19
Sensitivity	.74	.47	.92	.72
Specificity	.90	.98	.83	.95
Negative Predictive Power	.90	.82	.98	.95
Positive Predictive Power	.74	.89	.50	.74
Accurate Classification	.86	.83	.85	.92
Kappa	.64	.52	.56	.68

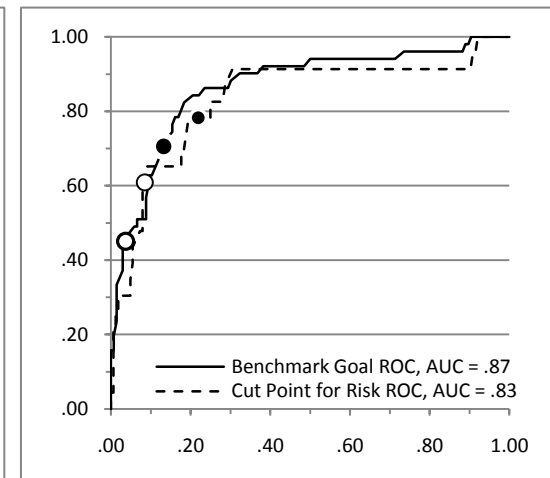
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3e	100	80	DORF Words Correct, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf3e Screening Decision:

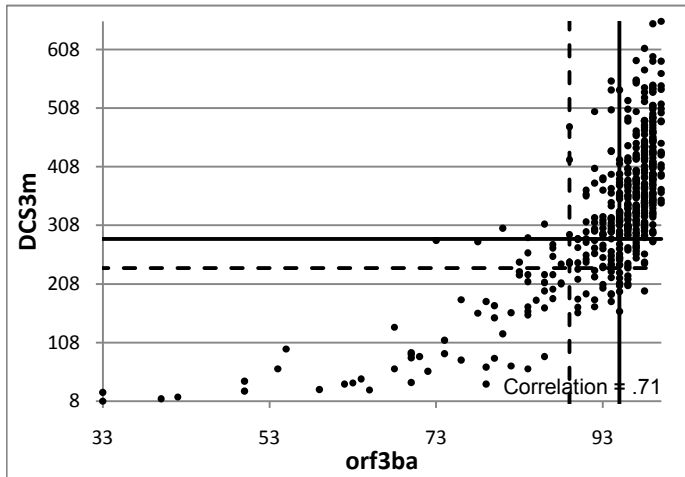
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr3e Outcome:					
At or Above Benchmark	5	13	118	136	73%
Below Benchmark	9	9	10	28	15%
Well Below Benchmark	14	4	5	23	12%
Marginal Total	28	26	133	187	
Marginal Percent	15%	14%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	18%	50%	89%		

At or Above Benchmark outcome

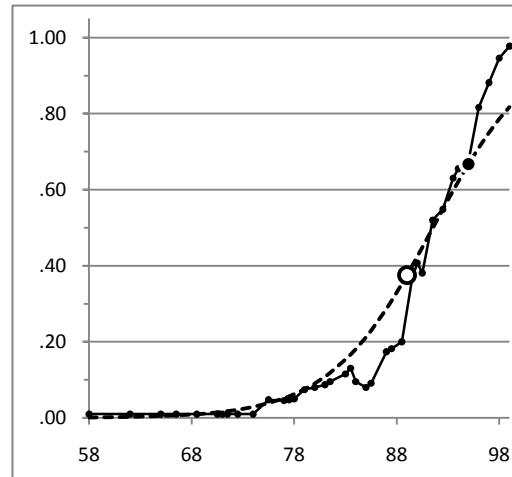
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	118	131	128	150
False Negative	15	28	5	9
True Positive	36	23	18	14
False Positive	18	5	36	14
Sensitivity	.71	.45	.78	.61
Specificity	.87	.96	.78	.91
Negative Predictive Power	.89	.82	.96	.94
Positive Predictive Power	.67	.82	.33	.50
Accurate Classification	.82	.82	.78	.88
Kappa	.56	.48	.36	.48

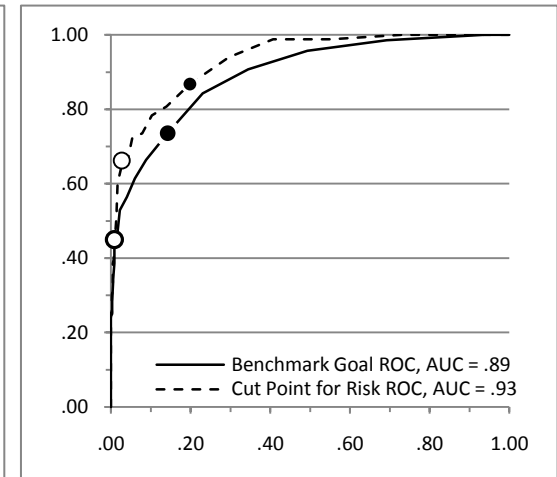
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3ba	95	89	DORF Accuracy, Grade 3, Beginning of Year
Outcome	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

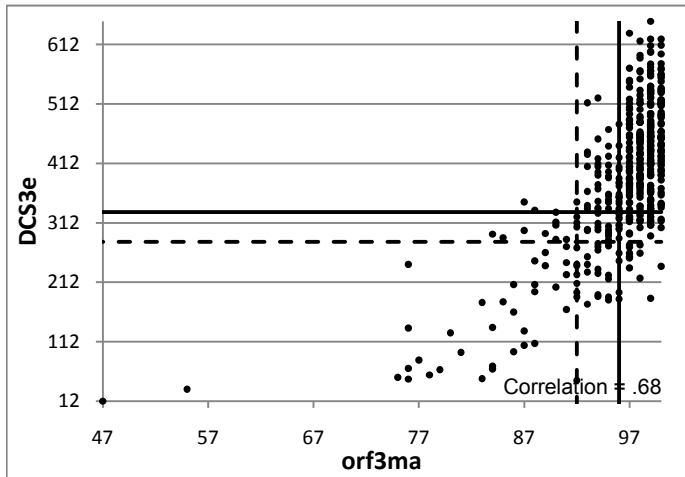
orf3ba Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3m Outcome:					
At or Above Benchmark	3	47	301	351	71%
Below Benchmark	8	23	26	57	12%
Well Below Benchmark	55	17	11	83	17%
Marginal Total	66	87	338	491	
Marginal Percent	13%	18%	69%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	5%	54%	89%		

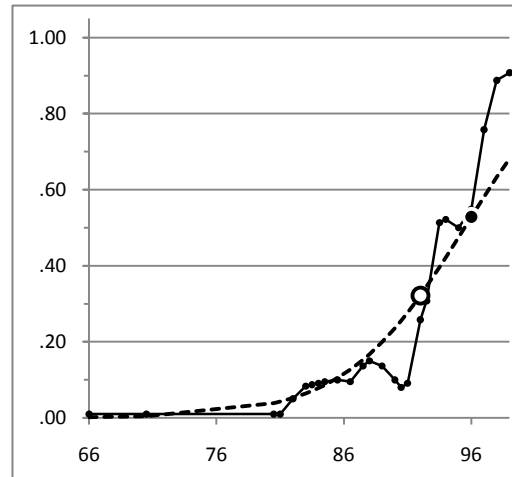
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	301	348	327	397
False Negative	37	77	11	28
True Positive	103	63	72	55
False Positive	50	3	81	11
Sensitivity	.74	.45	.87	.66
Specificity	.86	.99	.80	.97
Negative Predictive Power	.89	.82	.97	.93
Positive Predictive Power	.67	.95	.47	.83
Accurate Classification	.82	.84	.81	.92
Kappa	.58	.52	.50	.69

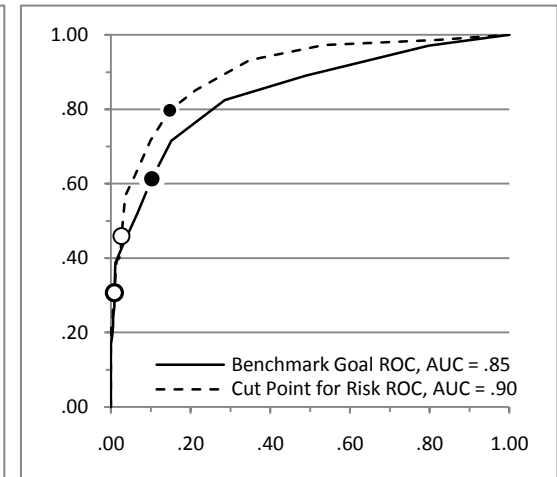
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3ma	96	92	DORF Accuracy, Grade 3, Middle of Year
Outcome	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

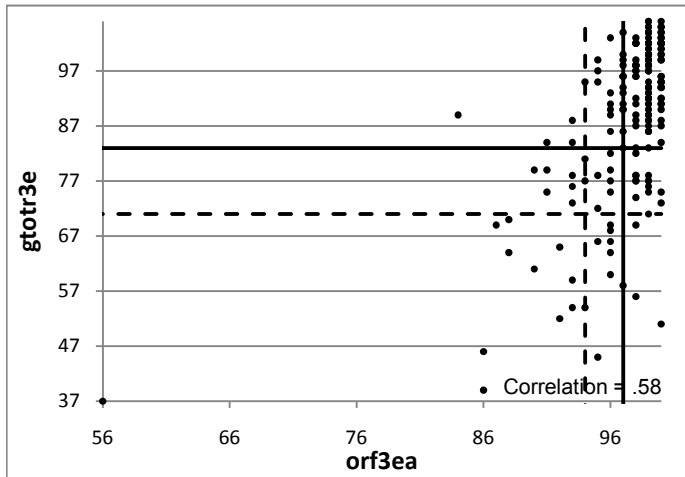
orf3ma Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3e Outcome:					
At or Above Benchmark	3	33	314	350	72%
Below Benchmark	8	17	38	63	13%
Well Below Benchmark	34	25	15	74	15%
Marginal Total	45	75	367	487	
Marginal Percent	9%	15%	75%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	44%	86%		

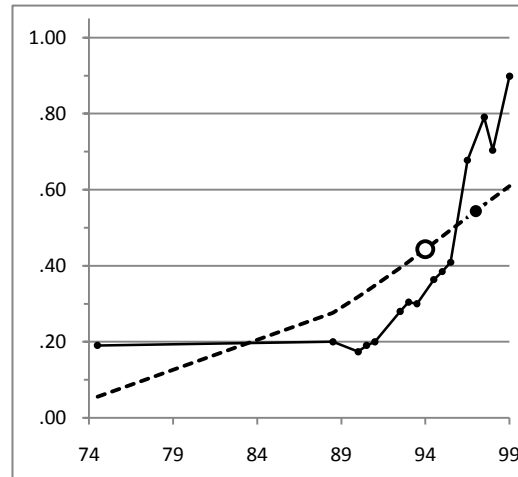
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	314	347	352	402
False Negative	53	95	15	40
True Positive	84	42	59	34
False Positive	36	3	61	11
Sensitivity	.61	.31	.80	.46
Specificity	.90	.99	.85	.97
Negative Predictive Power	.86	.79	.96	.91
Positive Predictive Power	.70	.93	.49	.76
Accurate Classification	.82	.80	.84	.90
Kappa	.53	.37	.52	.52

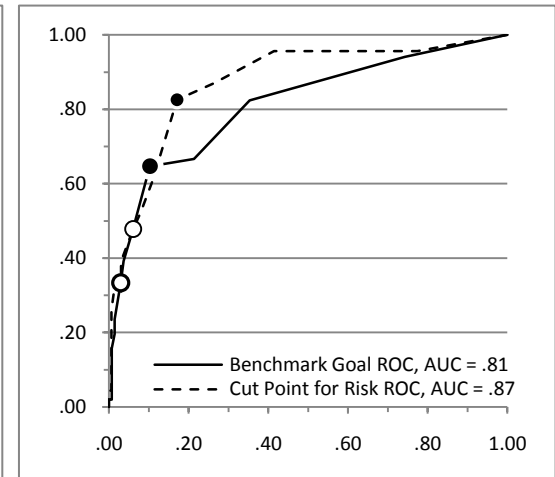
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf3ea	97	94	DORF Accuracy, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf3ea Screening Decision:

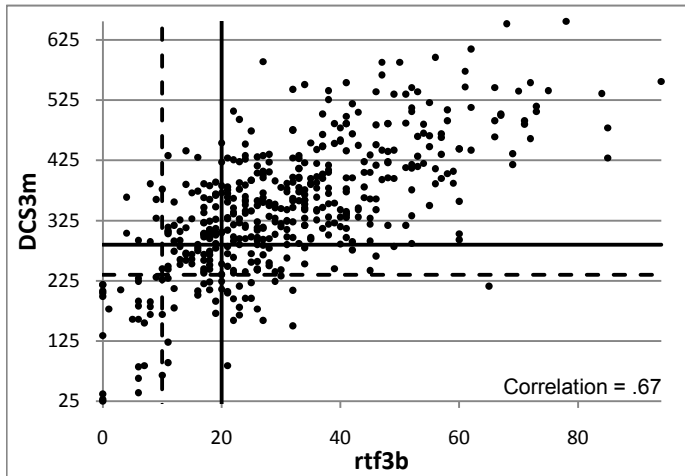
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr3e Outcome:					
At or Above Benchmark	4	10	122	136	73%
Below Benchmark	6	8	14	28	15%
Well Below Benchmark	11	8	4	23	12%
Marginal Total	21	26	140	187	
Marginal Percent	11%	14%	75%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	19%	38%	87%		

At or Above Benchmark outcome

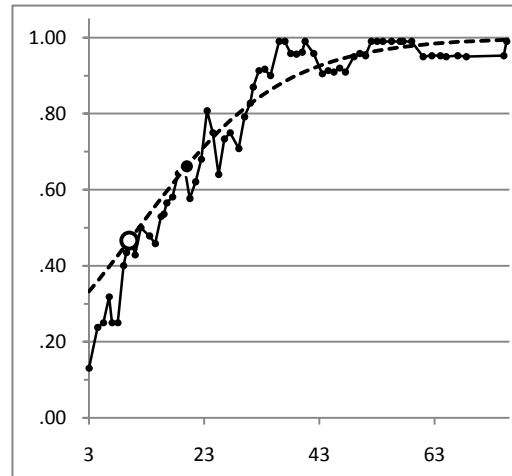
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	122	132	136	154
False Negative	18	34	4	12
True Positive	33	17	19	11
False Positive	14	4	28	10
Sensitivity	.65	.33	.83	.48
Specificity	.90	.97	.83	.94
Negative Predictive Power	.87	.80	.97	.93
Positive Predictive Power	.70	.81	.40	.52
Accurate Classification	.83	.80	.83	.88
Kappa	.56	.37	.45	.43

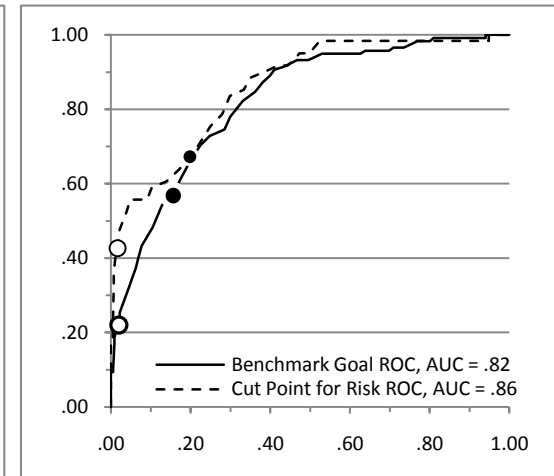
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf3b	20	10	DORF Retell, Grade 3, Beginning of Year
Outcome	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



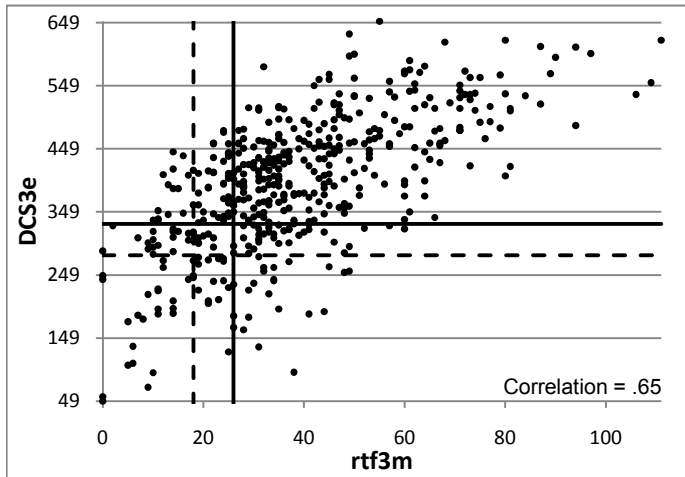
Receiver Operator Characteristic (ROC) curves.

rtf3b Screening Decision:

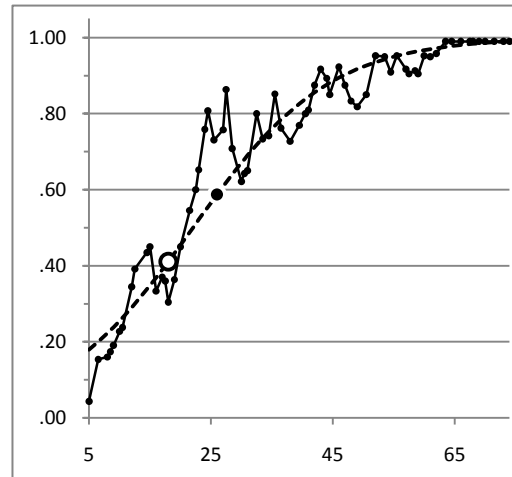
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3m Outcome:					
At or Above Benchmark	7	48	296	351	75%
Below Benchmark	0	26	31	57	12%
Well Below Benchmark	26	15	20	61	13%
Marginal Total	33	89	347	469	
Marginal Percent	7%	19%	74%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	21%	54%	85%		

	<u>At or Above Benchmark outcome</u>		<u>Well Below Benchmark outcome</u>	
	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	296	344	327	401
False Negative	51	92	20	35
True Positive	67	26	41	26
False Positive	55	7	81	7
Sensitivity	.57	.22	.67	.43
Specificity	.84	.98	.80	.98
Negative Predictive Power	.85	.79	.94	.92
Positive Predictive Power	.55	.79	.34	.79
Accurate Classification	.77	.79	.78	.91
Kappa	.41	.26	.33	.51

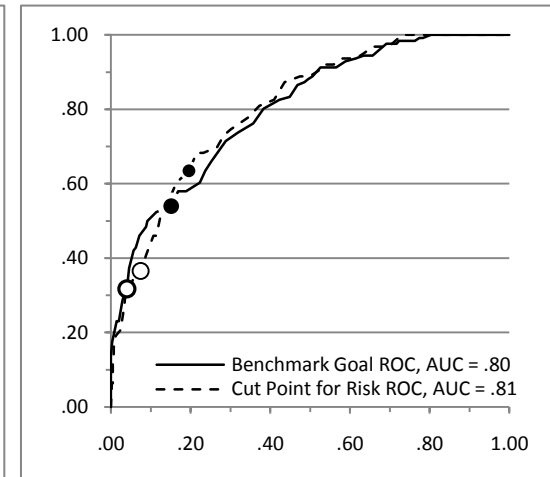
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf3m	26	18	DORF Retell, Grade 3, Middle of Year
Outcome	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

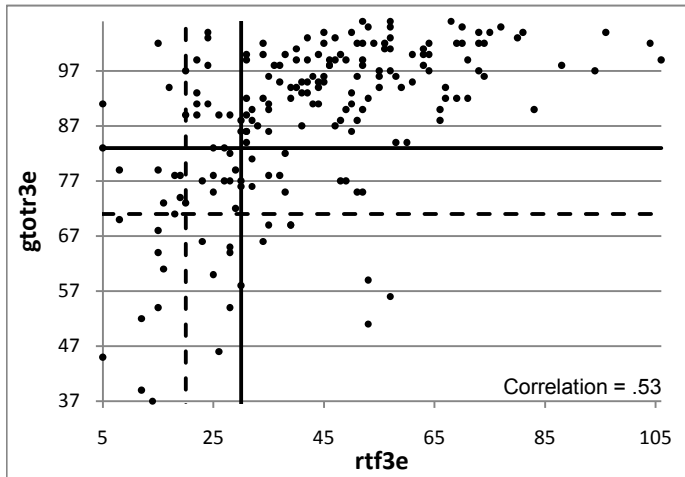
rtf3m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3e Outcome:					
At or Above Benchmark	14	39	297	350	74%
Below Benchmark	17	11	35	63	13%
Well Below Benchmark	23	17	23	63	13%
Marginal Total	54	67	355	476	
Marginal Percent	11%	14%	75%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	58%	84%		

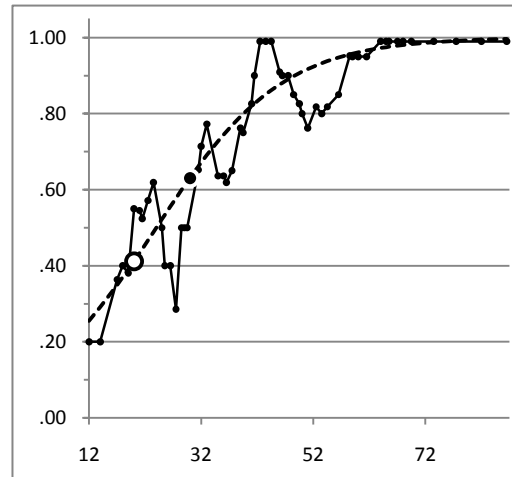
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	297	336	332	382
False Negative	58	86	23	40
True Positive	68	40	40	23
False Positive	53	14	81	31
Sensitivity	.54	.32	.63	.37
Specificity	.85	.96	.80	.92
Negative Predictive Power	.84	.80	.94	.91
Positive Predictive Power	.56	.74	.33	.43
Accurate Classification	.77	.79	.78	.85
Kappa	.39	.34	.32	.31

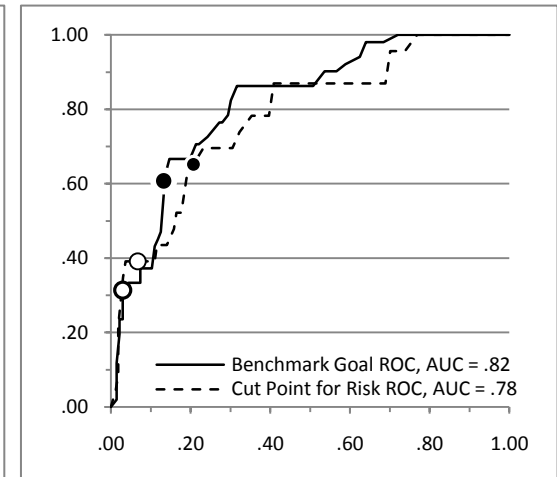
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf3e	30	20	DORF Retell, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

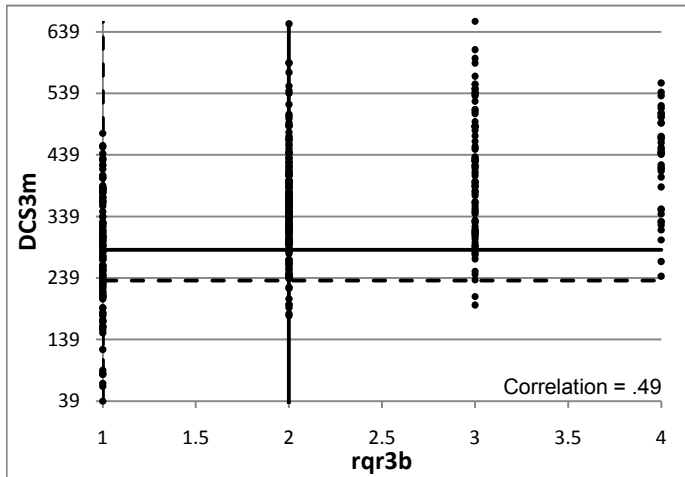
rtf3e Screening Decision:

gtotr3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	4	14	118	136	73%
Below Benchmark	7	9	12	28	15%
Well Below Benchmark	9	6	8	23	12%
Marginal Total	20	29	138	187	
Marginal Percent	11%	16%	74%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	48%	86%		

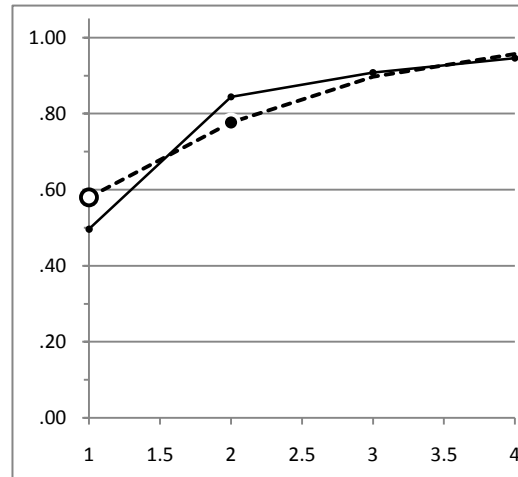
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	118	132	130	153
False Negative	20	35	8	14
True Positive	31	16	15	9
False Positive	18	4	34	11
Sensitivity	.61	.31	.65	.39
Specificity	.87	.97	.79	.93
Negative Predictive Power	.86	.79	.94	.92
Positive Predictive Power	.63	.80	.31	.45
Accurate Classification	.80	.79	.78	.87
Kappa	.48	.35	.30	.34

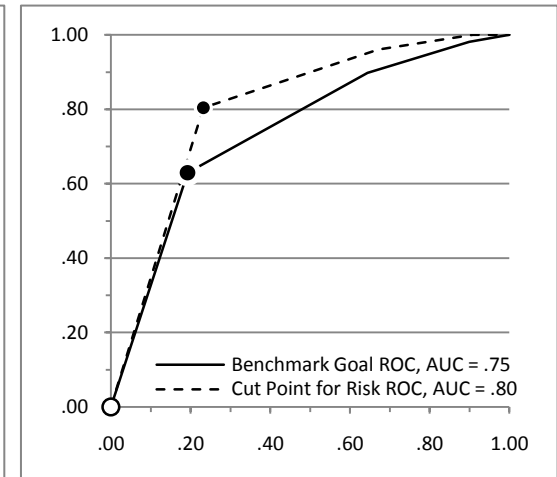
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr3b	2	1	DORF Retell Quality of Response, Grade 3, Beginning of Year
Outcome	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

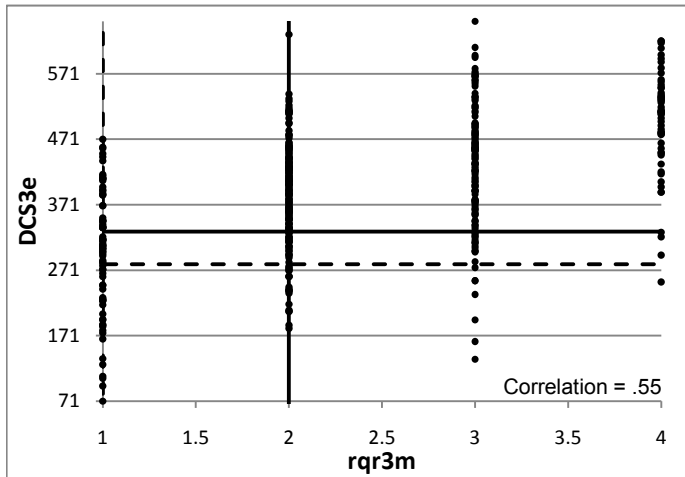
rqr3b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3m Outcome:					
At or Above Benchmark	0	67	281	348	76%
Below Benchmark	0	27	30	57	13%
Well Below Benchmark	0	41	10	51	11%
Marginal Total	0	135	321	456	
Marginal Percent	0%	30%	70%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	50%	88%		

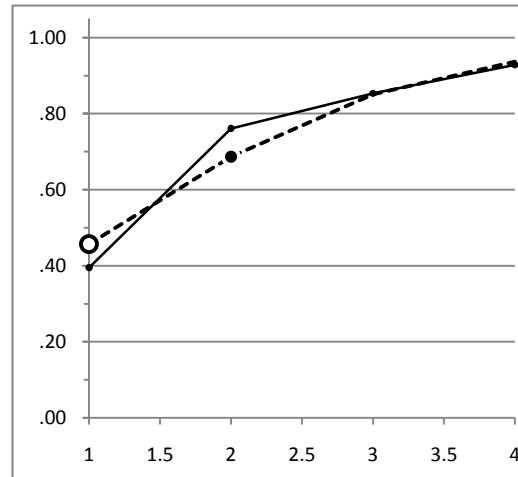
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	281	348	311	405
False Negative	40	108	10	51
True Positive	68	0	41	0
False Positive	67	0	94	0
Sensitivity	.63	.00	.80	.00
Specificity	.81	1.00	.77	1.00
Negative Predictive Power	.88	.76	.97	.89
Positive Predictive Power	.50	--	.30	--
Accurate Classification	.77	.76	.77	.89
Kappa	.40	.00	.33	.00

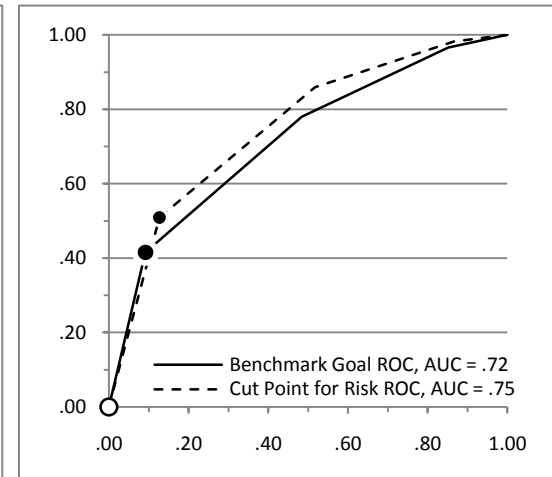
Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	rqr3m	2	1	DORF Retell Quality of Response, Grade 3, Middle of Year
	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

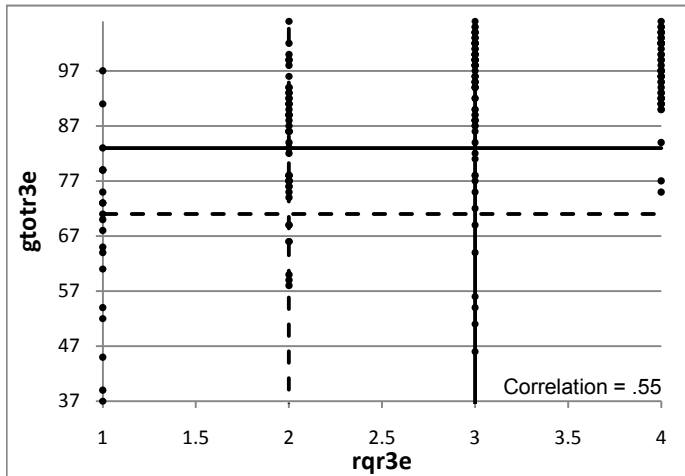
rqr3m Screening Decision:

DCS3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	0	32	317	349	75%
Below Benchmark	0	20	41	61	13%
Well Below Benchmark	0	29	28	57	12%
Marginal Total	0	81	386	467	
Marginal Percent	0%	17%	83%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	40%	82%		

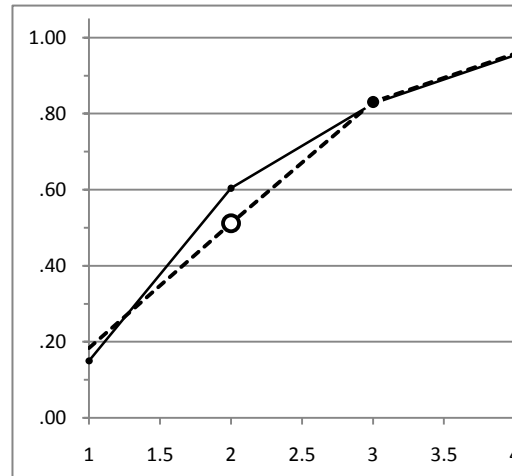
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	317	349	358	410
False Negative	69	118	28	57
True Positive	49	0	29	0
False Positive	32	0	52	0
Sensitivity	.42	.00	.51	.00
Specificity	.91	1.00	.87	1.00
Negative Predictive Power	.82	.75	.93	.88
Positive Predictive Power	.60	--	.36	--
Accurate Classification	.78	.75	.83	.88
Kappa	.36	.00	.32	.00

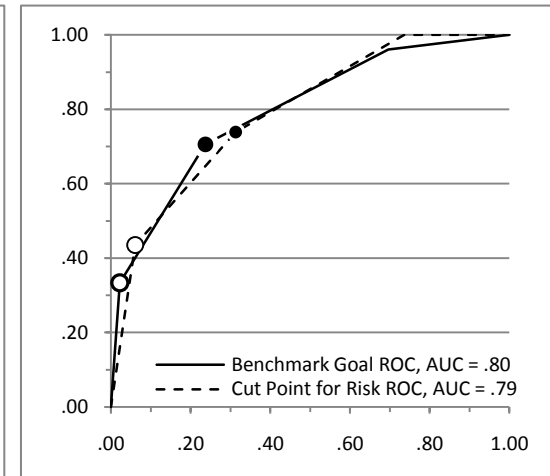
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr3e	3	2	DORF Retell Quality of Response, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

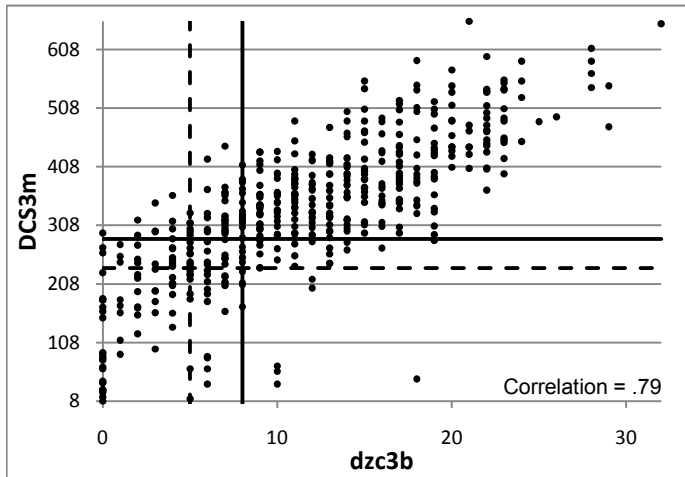
rqr3e Screening Decision:

gtotr3e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	3	29	103	135	73%
Below Benchmark	7	12	9	28	15%
Well Below Benchmark	10	7	6	23	12%
Marginal Total	20	48	118	186	
Marginal Percent	11%	26%	63%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	15%	60%	87%		

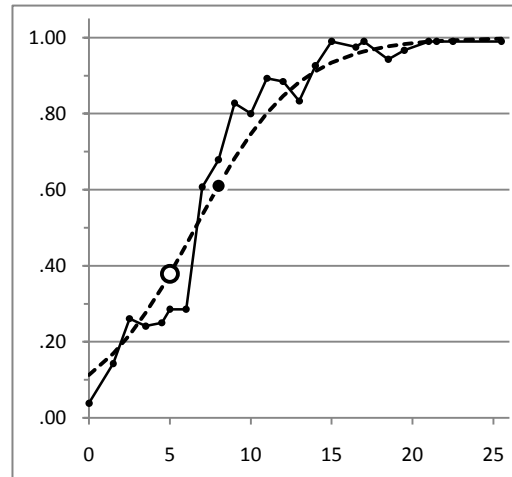
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	103	132	112	153
False Negative	15	34	6	13
True Positive	36	17	17	10
False Positive	32	3	51	10
Sensitivity	.71	.33	.74	.43
Specificity	.76	.98	.69	.94
Negative Predictive Power	.87	.80	.95	.92
Positive Predictive Power	.53	.85	.25	.50
Accurate Classification	.75	.80	.69	.88
Kappa	.42	.38	.23	.40

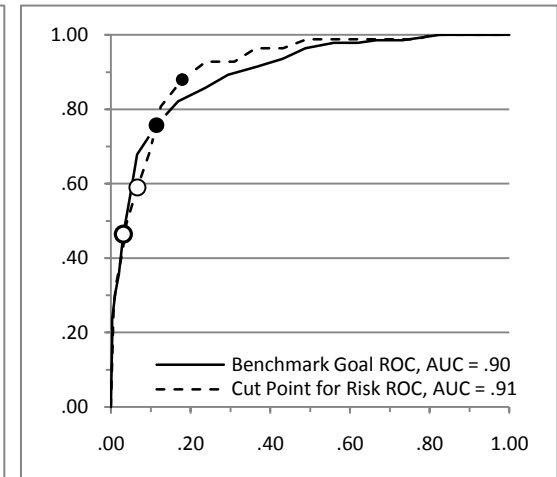
Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	dzc3b	8	5	Daze Adjusted Score, Grade 3, Beginning of Year
	Criterion	DCS3m	285	235	DIBELS Composite Score, Grade 3, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

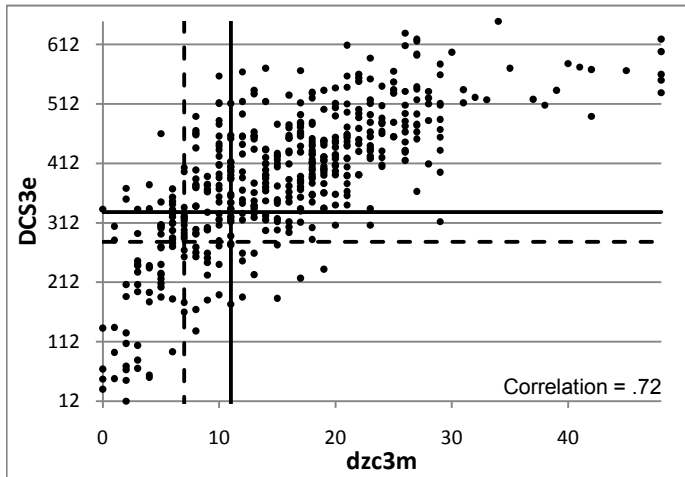
dzc3b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3m Outcome:					
At or Above Benchmark	11	29	310	350	71%
Below Benchmark	16	17	24	57	12%
Well Below Benchmark	49	24	10	83	17%
Marginal Total	76	70	344	490	
Marginal Percent	16%	14%	70%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	41%	90%		

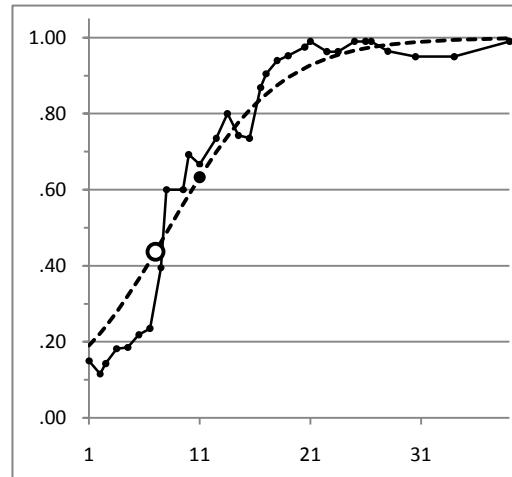
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	310	339	334	380
False Negative	34	75	10	34
True Positive	106	65	73	49
False Positive	40	11	73	27
Sensitivity	.76	.46	.88	.59
Specificity	.89	.97	.82	.93
Negative Predictive Power	.90	.82	.97	.92
Positive Predictive Power	.73	.86	.50	.64
Accurate Classification	.85	.82	.83	.88
Kappa	.63	.50	.54	.54

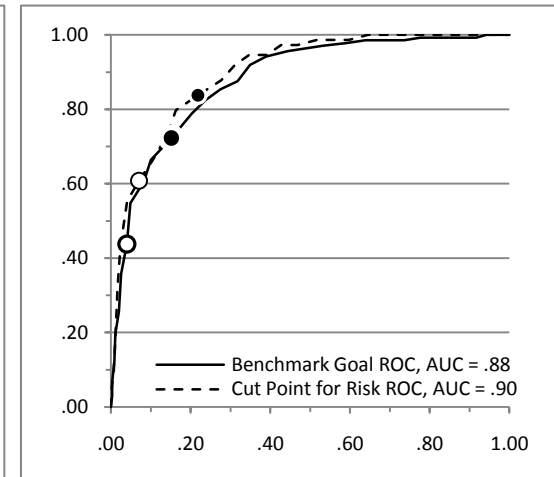
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc3m	11	7	Daze Adjusted Score, Grade 3, Middle of Year
Outcome	Criterion	DCS3e	330	280	DIBELS Composite Score, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

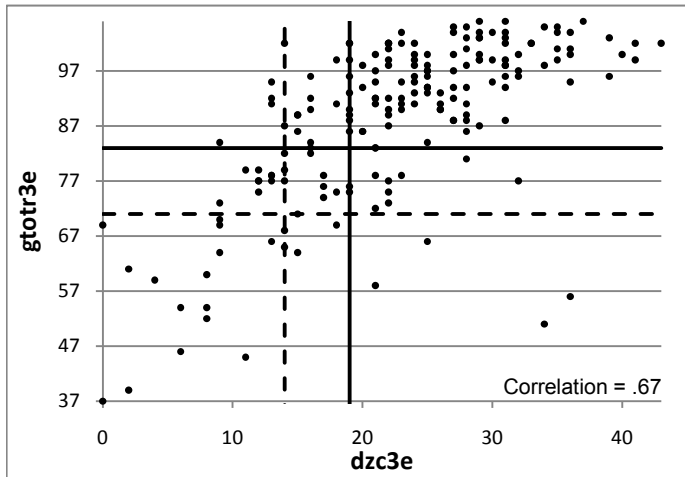
dzc3m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS3e Outcome:					
At or Above Benchmark	14	39	296	349	72%
Below Benchmark	15	22	26	63	13%
Well Below Benchmark	45	17	12	74	15%
Marginal Total	74	78	334	486	
Marginal Percent	15%	16%	69%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	19%	50%	89%		

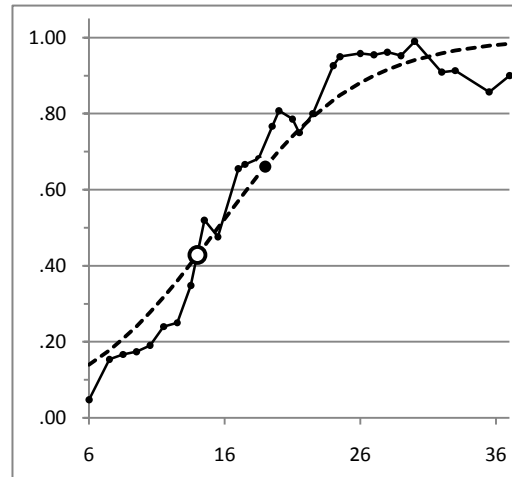
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	296	335	322	383
False Negative	38	77	12	29
True Positive	99	60	62	45
False Positive	53	14	90	29
Sensitivity	.72	.44	.84	.61
Specificity	.85	.96	.78	.93
Negative Predictive Power	.89	.81	.96	.93
Positive Predictive Power	.65	.81	.41	.61
Accurate Classification	.81	.81	.79	.88
Kappa	.55	.46	.43	.54

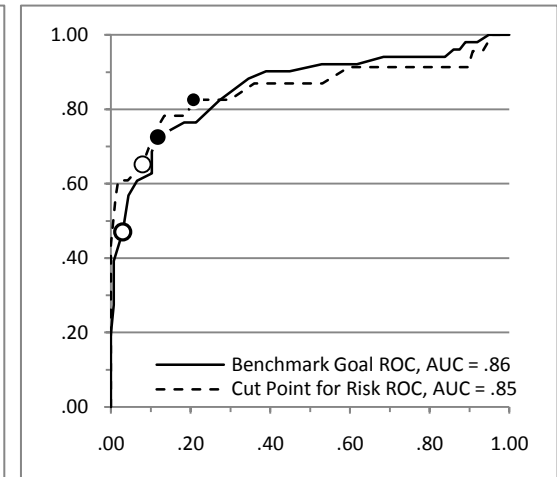
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc3e	19	14	Daze Adjusted Score, Grade 3, End of Year
Outcome	Criterion	gtotr3e	83	71	GRADE Total Test, Grade 3, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

dzc3e Screening Decision:

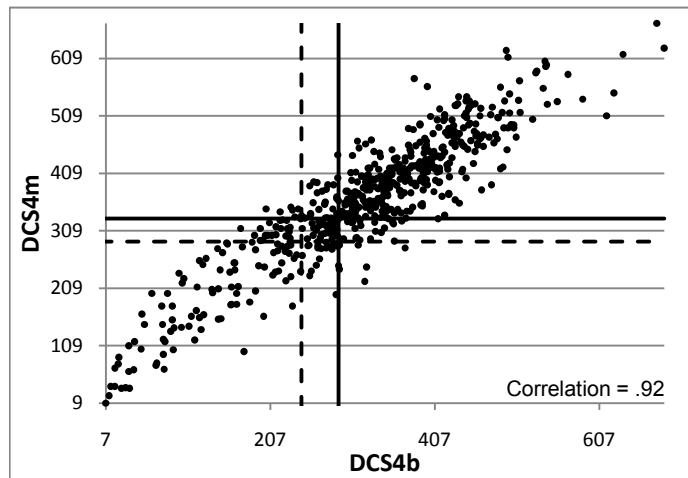
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr3e Outcome:					
At or Above Benchmark	4	12	120	136	73%
Below Benchmark	9	9	10	28	15%
Well Below Benchmark	15	4	4	23	12%
Marginal Total	28	25	134	187	
Marginal Percent	15%	13%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	48%	90%		

At or Above Benchmark outcome

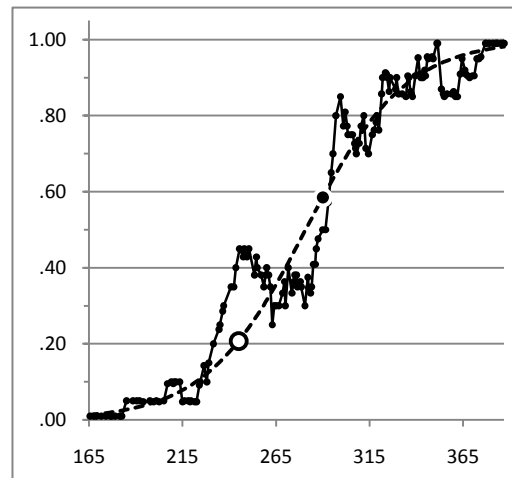
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	120	132	130	151
False Negative	14	27	4	8
True Positive	37	24	19	15
False Positive	16	4	34	13
Sensitivity	.73	.47	.83	.65
Specificity	.88	.97	.79	.92
Negative Predictive Power	.90	.83	.97	.95
Positive Predictive Power	.70	.86	.36	.54
Accurate Classification	.84	.83	.80	.89
Kappa	.60	.51	.40	.52

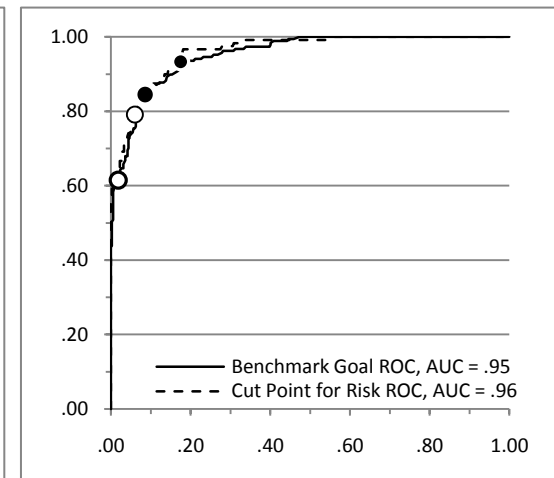
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS4b	290	245	DIBELS Composite Score, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

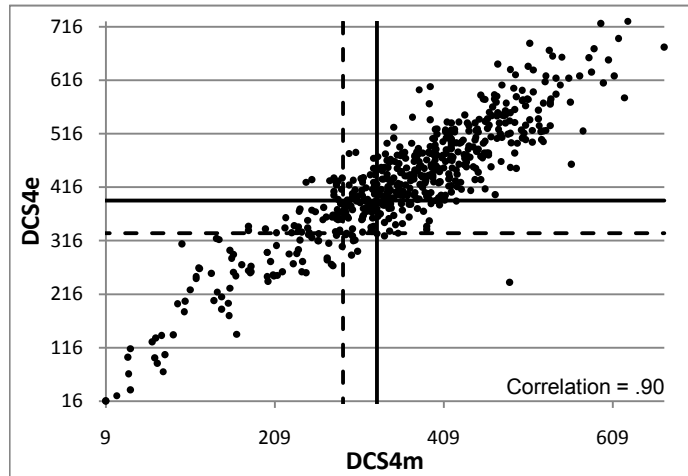
DCS4b Screening Decision:

DCS4m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	7	26	351	384	67%
Below Benchmark	20	26	21	67	12%
Well Below Benchmark	95	17	8	120	21%
Marginal Total	122	69	380	571	
Marginal Percent	21%	12%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	38%	92%		

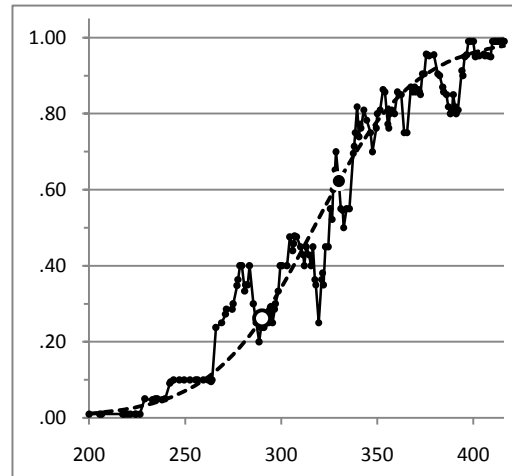
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	351	377	372	424
False Negative	29	72	8	25
True Positive	158	115	112	95
False Positive	33	7	79	27
Sensitivity	.84	.61	.93	.79
specificity	.91	.98	.82	.94
Negative Predictive Power	.92	.84	.98	.94
Positive Predictive Power	.83	.94	.59	.78
Accurate Classification	.89	.86	.85	.91
Kappa	.75	.66	.62	.73

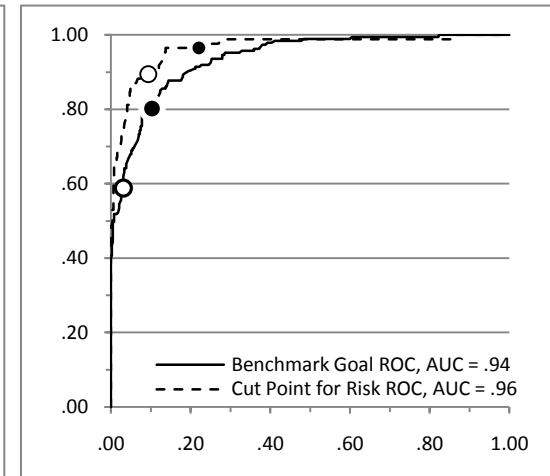
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year
Outcome	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

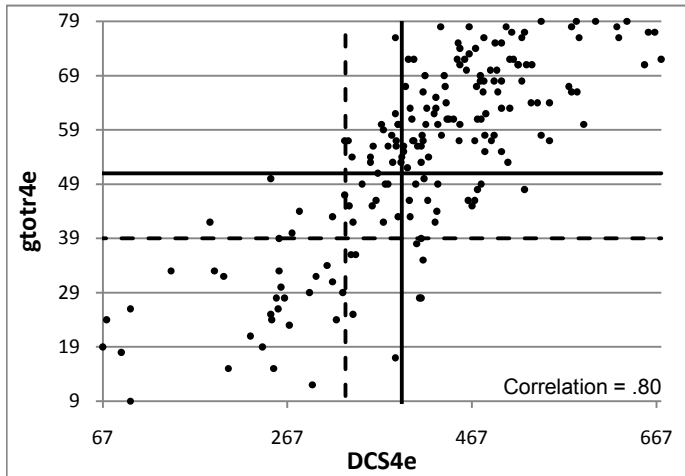
DCS4m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	12	28	347	387	67%
Below Benchmark	34	34	34	102	18%
Well Below Benchmark	76	6	3	85	15%
Marginal Total	122	68	384	574	
Marginal Percent	21%	12%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	41%	90%		

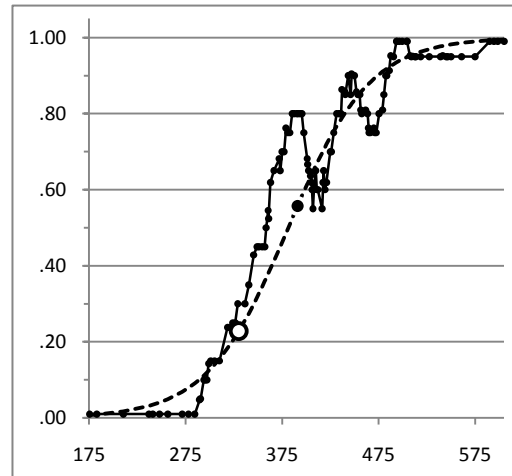
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	347	375	381	443
False Negative	37	77	3	9
True Positive	150	110	82	76
False Positive	40	12	108	46
Sensitivity	.80	.59	.96	.89
specificity	.90	.97	.78	.91
Negative Predictive Power	.90	.83	.99	.98
Positive Predictive Power	.79	.90	.43	.62
Accurate Classification	.87	.84	.81	.90
Kappa	.70	.61	.49	.68

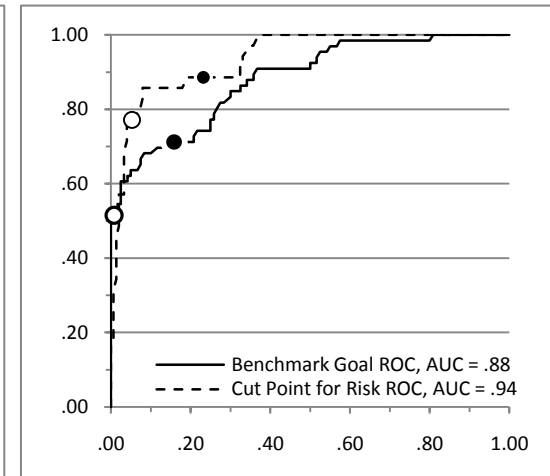
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

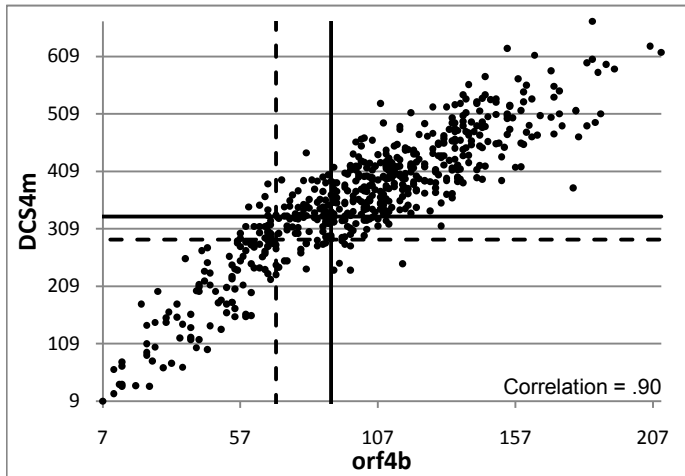
DCS4e Screening Decision:

gtotr4e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	1	18	101	120	65%
Below Benchmark	7	9	15	31	17%
Well Below Benchmark	27	4	4	35	19%
Marginal Total	35	31	120	186	
Marginal Percent	19%	17%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	3%	58%	84%		

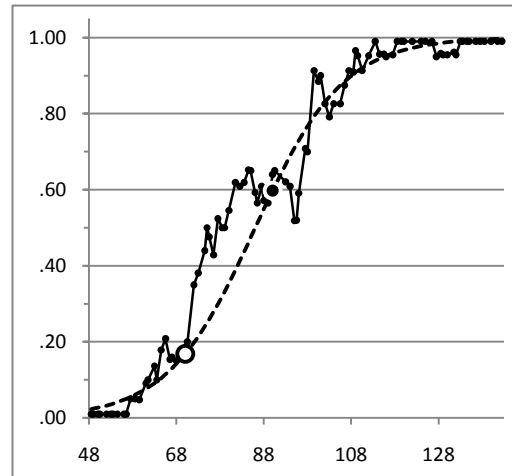
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	101	119	116	143
False Negative	19	32	4	8
True Positive	47	34	31	27
False Positive	19	1	35	8
Sensitivity	.71	.52	.89	.77
specificity	.84	.99	.77	.95
Negative Predictive Power	.84	.79	.97	.95
Positive Predictive Power	.71	.97	.47	.77
Accurate Classification	.80	.82	.79	.91
Kappa	.55	.57	.49	.72

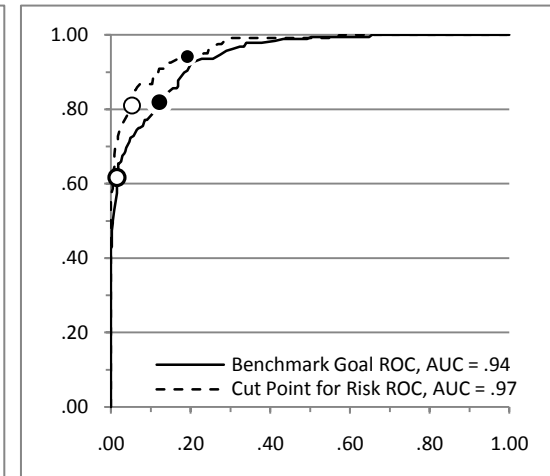
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4b	90	70	DORF Words Correct, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf4b Screening Decision:

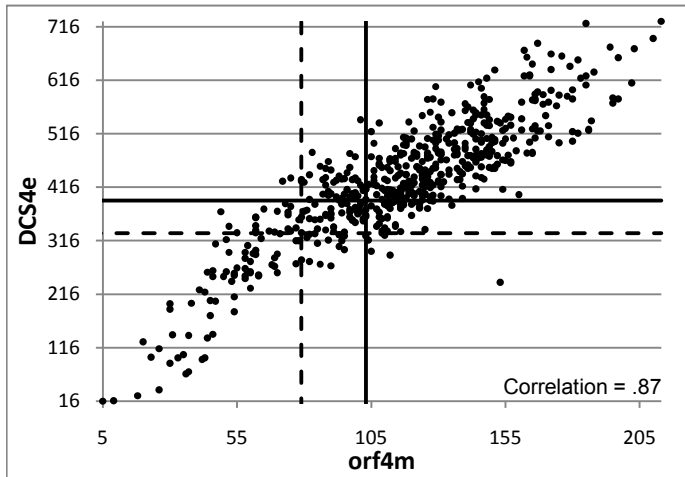
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4m Outcome:					
At or Above Benchmark	6	41	339	386	67%
Below Benchmark	18	22	27	67	12%
Well Below Benchmark	98	16	7	121	21%
Marginal Total	122	79	373	574	
Marginal Percent	21%	14%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	5%	52%	91%		

At or Above Benchmark outcome

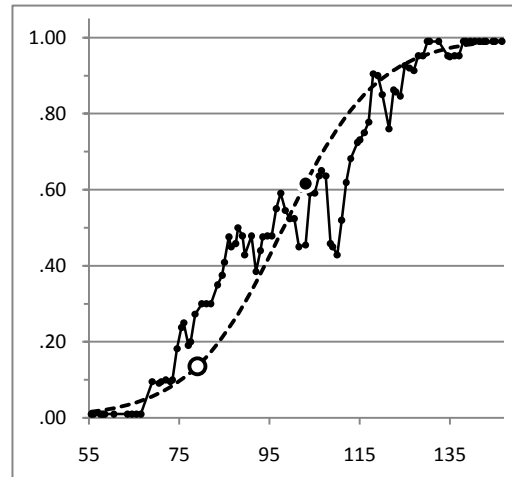
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	339	380	366	429
False Negative	34	72	7	23
True Positive	154	116	114	98
False Positive	47	6	87	24
Sensitivity	.82	.62	.94	.81
specificity	.88	.98	.81	.95
Negative Predictive Power	.91	.84	.98	.95
Positive Predictive Power	.77	.95	.57	.80
Accurate Classification	.86	.86	.84	.92
Kappa	.69	.66	.60	.75

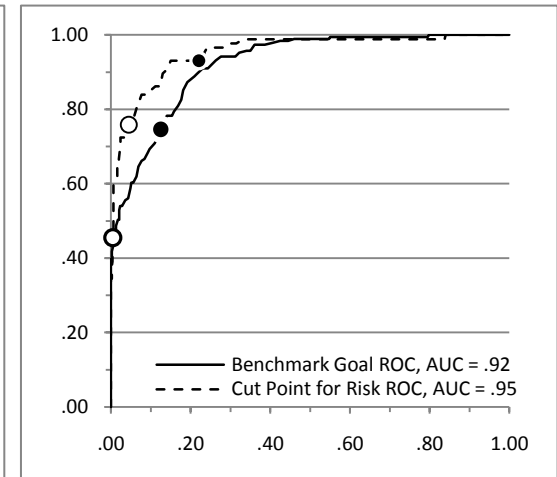
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4m	103	79	DORF Words Correct, Grade 4, Middle of Year
Outcome	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

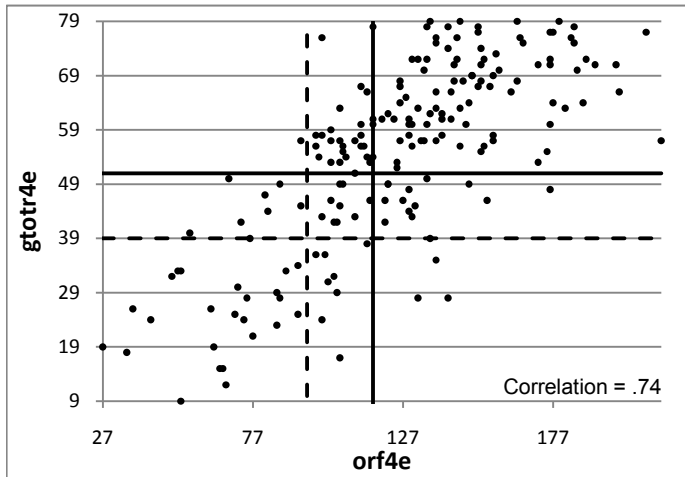
orf4m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	2	47	342	391	67%
Below Benchmark	20	40	42	102	18%
Well Below Benchmark	66	15	6	87	15%
Marginal Total	88	102	390	580	
Marginal Percent	15%	18%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	2%	46%	88%		

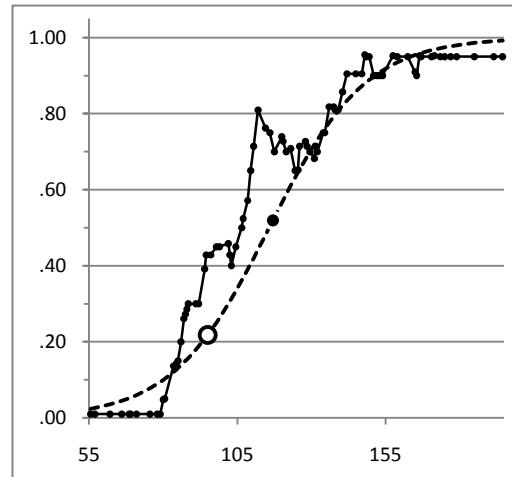
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	342	389	384	471
False Negative	48	103	6	21
True Positive	141	86	81	66
False Positive	49	2	109	22
Sensitivity	.75	.46	.93	.76
specificity	.87	.99	.78	.96
Negative Predictive Power	.88	.79	.98	.96
Positive Predictive Power	.74	.98	.43	.75
Accurate Classification	.83	.82	.80	.93
Kappa	.62	.52	.48	.71

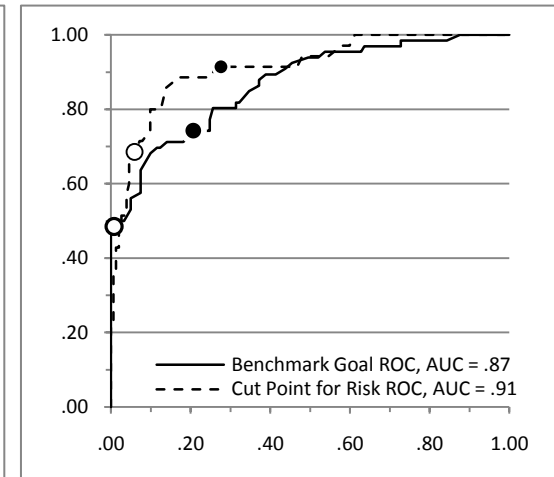
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4e	117	95	DORF Words Correct, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

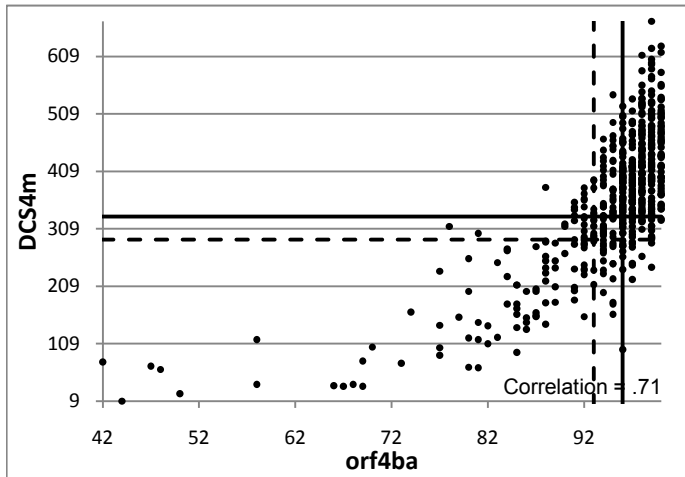
orf4e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr4e Outcome:					
At or Above Benchmark	1	24	96	121	65%
Below Benchmark	8	9	14	31	17%
Well Below Benchmark	24	8	3	35	19%
Marginal Total	33	41	113	187	
Marginal Percent	18%	22%	60%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	3%	59%	85%		

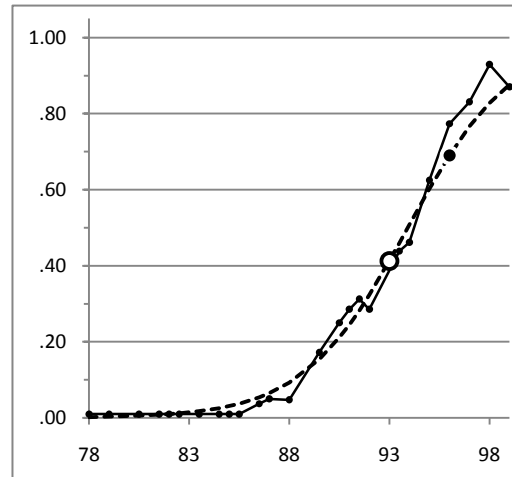
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	96	120	110	143
False Negative	17	34	3	11
True Positive	49	32	32	24
False Positive	25	1	42	9
Sensitivity	.74	.48	.91	.69
specificity	.79	.99	.72	.94
Negative Predictive Power	.85	.78	.97	.93
Positive Predictive Power	.66	.97	.43	.73
Accurate Classification	.78	.81	.76	.89
Kappa	.52	.54	.45	.64

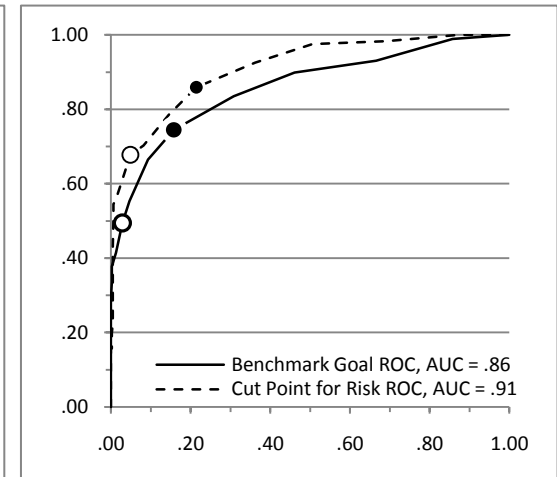
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4ba	96	93	DORF Accuracy, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

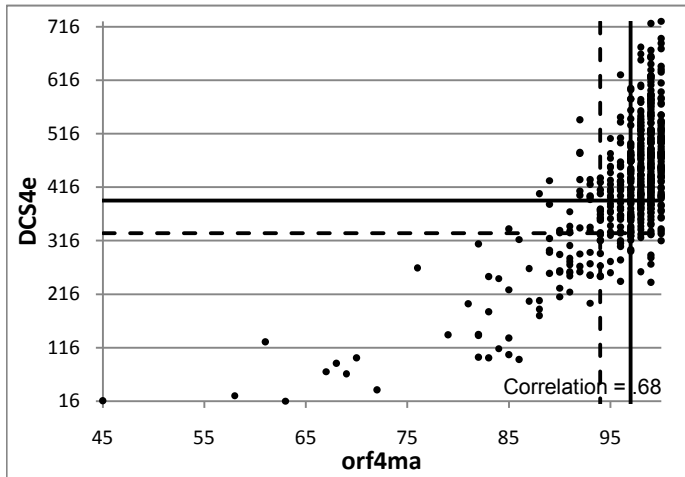
orf4ba Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4m Outcome:					
At or Above Benchmark	11	50	325	386	67%
Below Benchmark	11	25	31	67	12%
Well Below Benchmark	82	22	17	121	21%
Marginal Total	104	97	373	574	
Marginal Percent	18%	17%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	11%	52%	87%		

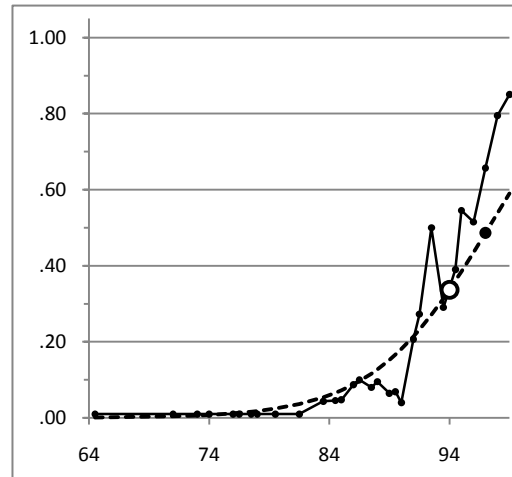
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	325	375	356	431
False Negative	48	95	17	39
True Positive	140	93	104	82
False Positive	61	11	97	22
Sensitivity	.74	.49	.86	.68
specificity	.84	.97	.79	.95
Negative Predictive Power	.87	.80	.95	.92
Positive Predictive Power	.70	.89	.52	.79
Accurate Classification	.81	.82	.80	.89
Kappa	.58	.53	.52	.66

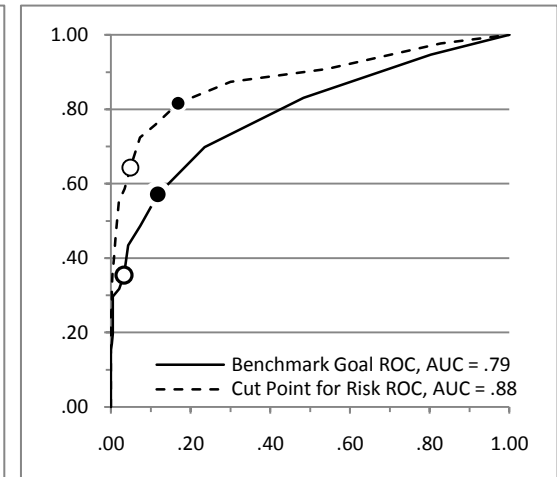
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4ma	97	94	DORF Accuracy, Grade 4, Middle of Year
Outcome	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

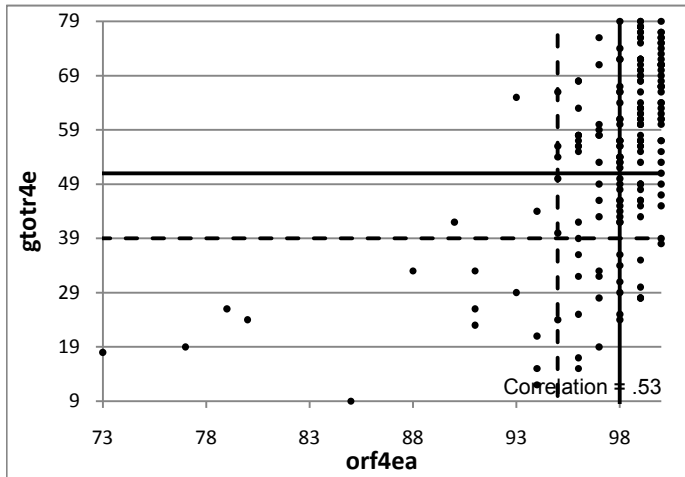
orf4ma Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	13	33	345	391	67%
Below Benchmark	11	26	65	102	18%
Well Below Benchmark	56	15	16	87	15%
Marginal Total	80	74	426	580	
Marginal Percent	14%	13%	73%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	16%	45%	81%		

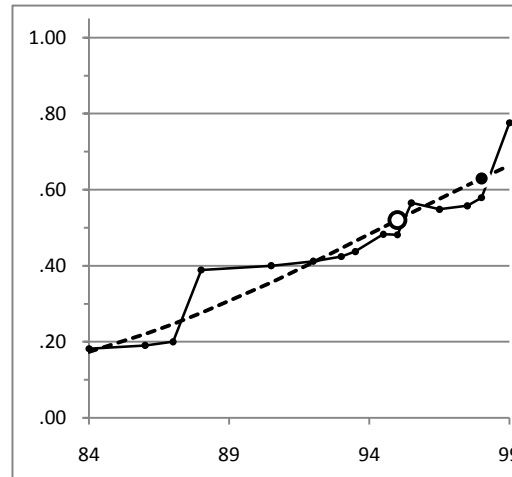
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	345	378	410	469
False Negative	81	122	16	31
True Positive	108	67	71	56
False Positive	46	13	83	24
Sensitivity	.57	.35	.82	.64
specificity	.88	.97	.83	.95
Negative Predictive Power	.81	.76	.96	.94
Positive Predictive Power	.70	.84	.46	.70
Accurate Classification	.78	.77	.83	.91
Kappa	.48	.38	.49	.62

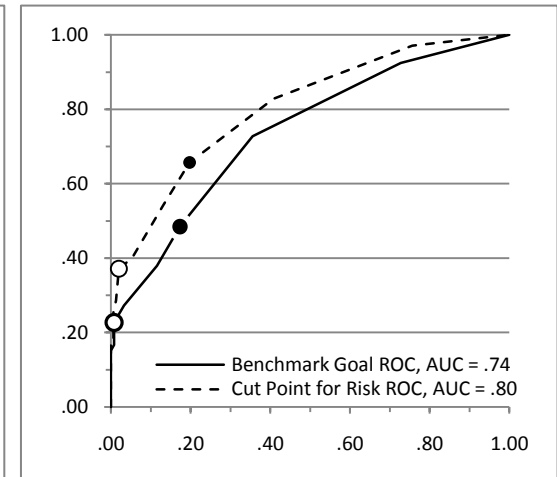
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf4ea	98	95	DORF Accuracy, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

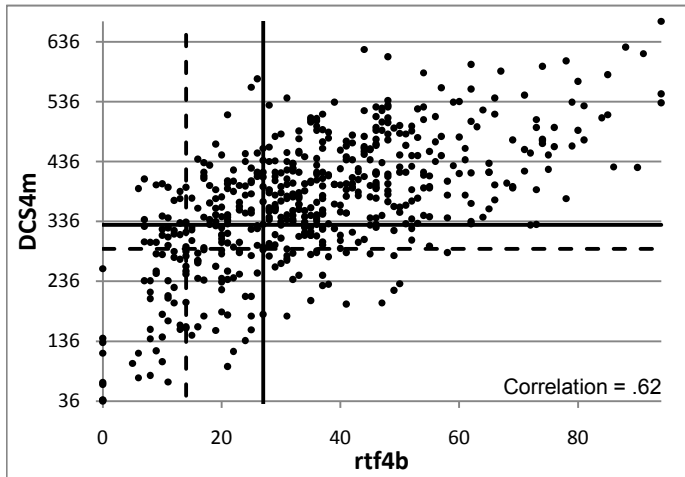
orf4ea Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr4e Outcome:					
At or Above Benchmark	1	20	100	121	65%
Below Benchmark	2	7	22	31	17%
Well Below Benchmark	13	10	12	35	19%
Marginal Total	16	37	134	187	
Marginal Percent	9%	20%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	54%	75%		

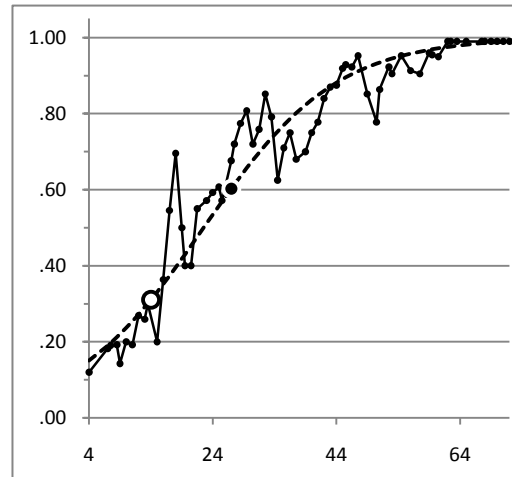
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	100	120	122	149
False Negative	34	51	12	22
True Positive	32	15	23	13
False Positive	21	1	30	3
Sensitivity	.48	.23	.66	.37
specificity	.83	.99	.80	.98
Negative Predictive Power	.75	.70	.91	.87
Positive Predictive Power	.60	.94	.43	.81
Accurate Classification	.71	.72	.78	.87
Kappa	.33	.26	.38	.44

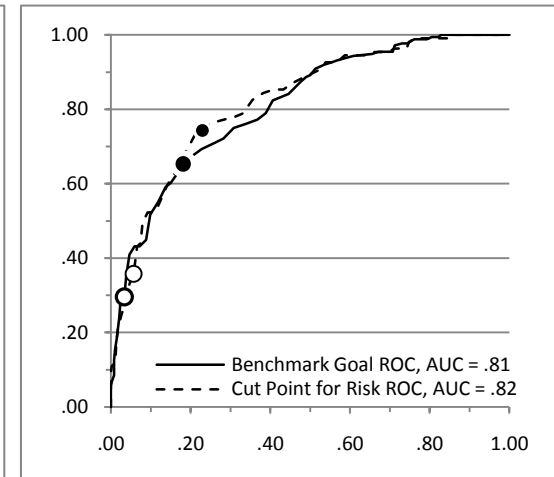
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf4b	27	14	DORF Retell, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

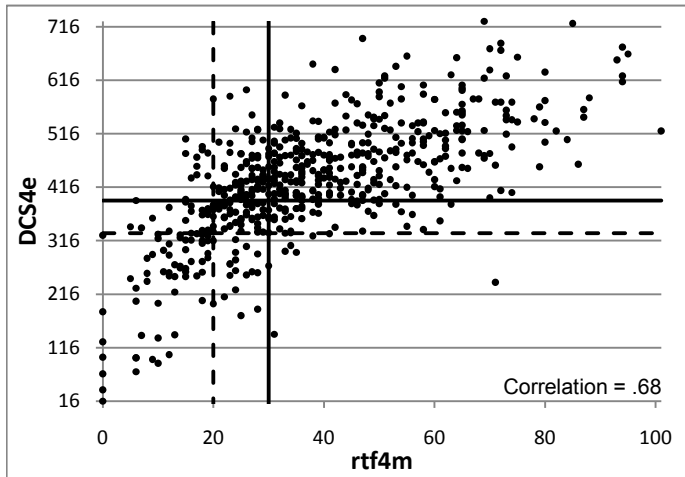
rtf4b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4m Outcome:					
At or Above Benchmark	13	57	316	386	69%
Below Benchmark	13	21	33	67	12%
Well Below Benchmark	39	42	28	109	19%
Marginal Total	65	120	377	562	
Marginal Percent	12%	21%	67%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	48%	84%		

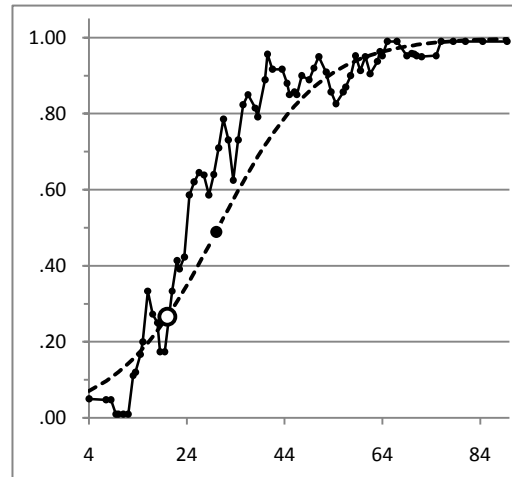
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	316	373	349	427
False Negative	61	124	28	70
True Positive	115	52	81	39
False Positive	70	13	104	26
Sensitivity	.65	.30	.74	.36
specificity	.82	.97	.77	.94
Negative Predictive Power	.84	.75	.93	.86
Positive Predictive Power	.62	.80	.44	.60
Accurate Classification	.77	.76	.77	.83
Kappa	.47	.32	.41	.35

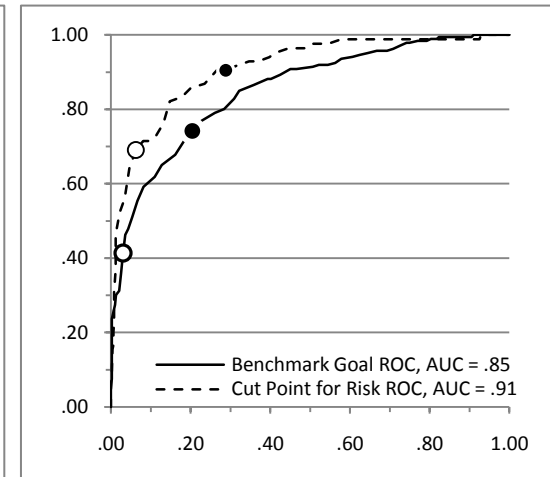
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf4m	30	20	DORF Retell, Grade 4, Middle of Year
Outcome	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

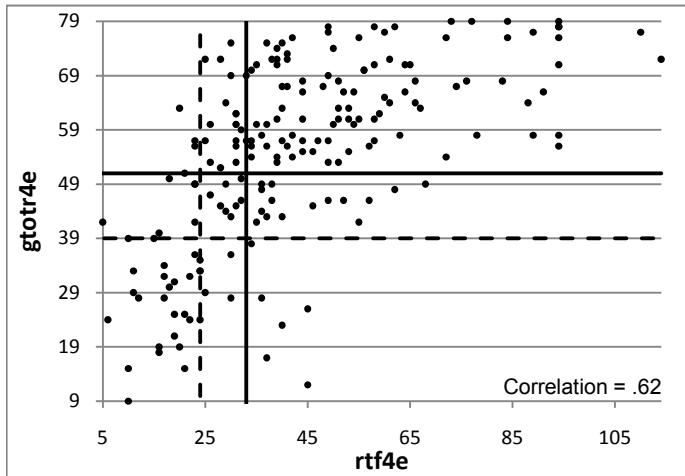
rtf4m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	12	68	311	391	68%
Below Benchmark	19	43	40	102	18%
Well Below Benchmark	58	18	8	84	15%
Marginal Total	89	129	359	577	
Marginal Percent	15%	22%	62%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	13%	53%	87%		

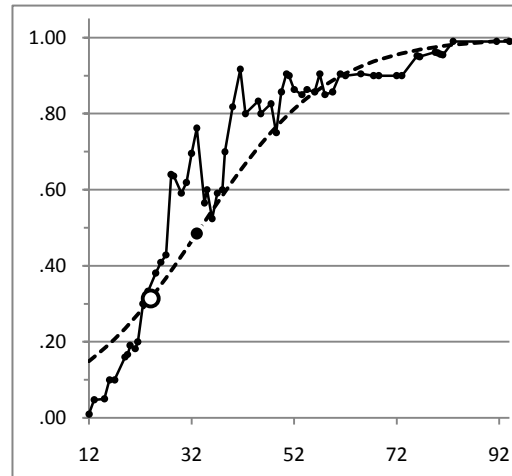
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	311	379	351	462
False Negative	48	109	8	26
True Positive	138	77	76	58
False Positive	80	12	142	31
Sensitivity	.74	.41	.90	.69
specificity	.80	.97	.71	.94
Negative Predictive Power	.87	.78	.98	.95
Positive Predictive Power	.63	.87	.35	.65
Accurate Classification	.78	.79	.74	.90
Kappa	.51	.44	.37	.61

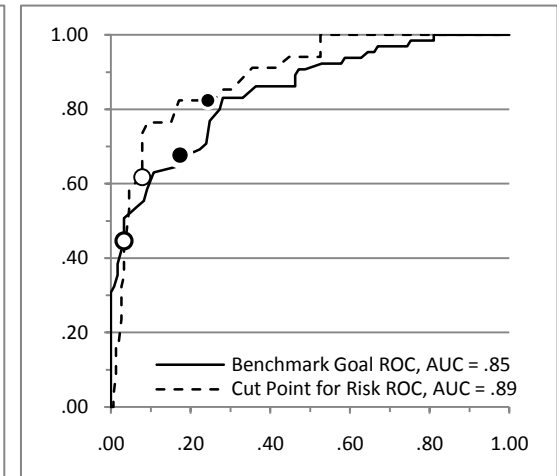
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf4e	33	24	DORF Retell, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

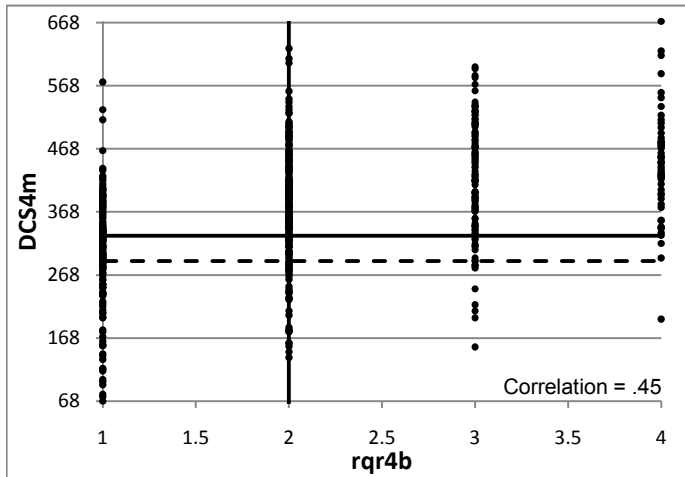
rtf4e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr4e Outcome:					
At or Above Benchmark	4	17	100	121	65%
Below Benchmark	8	8	15	31	17%
Well Below Benchmark	21	7	6	34	18%
Marginal Total	33	32	121	186	
Marginal Percent	18%	17%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	53%	83%		

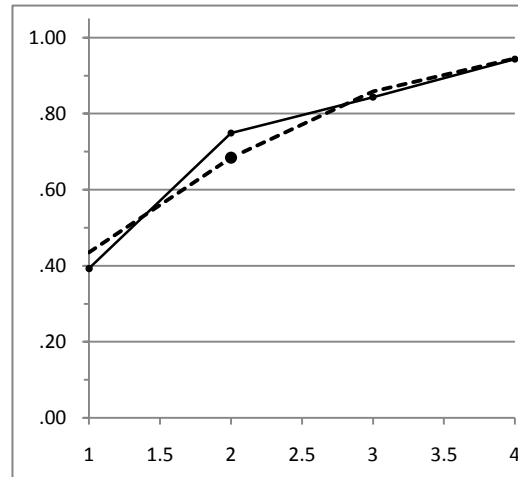
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	100	117	115	140
False Negative	21	36	6	13
True Positive	44	29	28	21
False Positive	21	4	37	12
Sensitivity	.68	.45	.82	.62
specificity	.83	.97	.76	.92
Negative Predictive Power	.83	.76	.95	.92
Positive Predictive Power	.68	.88	.43	.64
Accurate Classification	.77	.78	.77	.87
Kappa	.50	.47	.43	.54

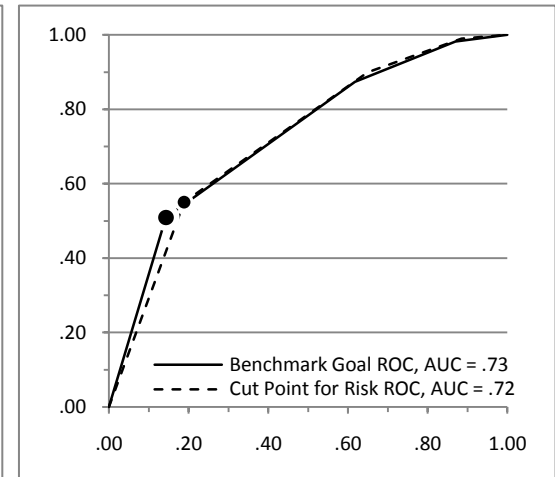
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr4b	2	0	DORF Retell Quality of Response, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

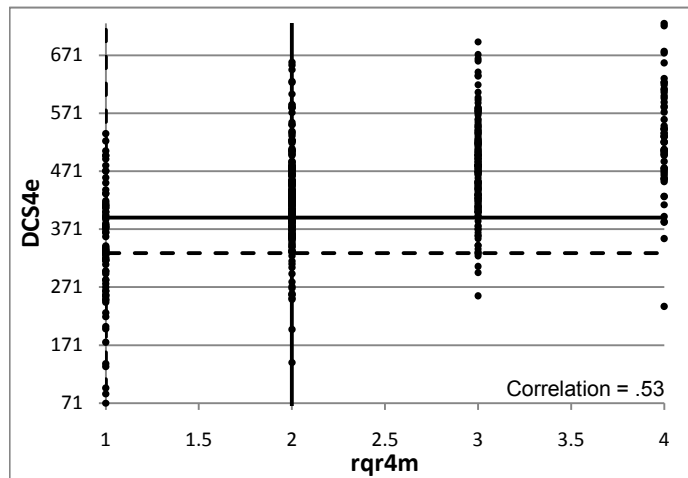
rqr4b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4m Outcome:					
At or Above Benchmark	0	55	329	384	70%
Below Benchmark	0	30	37	67	12%
Well Below Benchmark	0	55	45	100	18%
Marginal Total	0	140	411	551	
Marginal Percent	0%	25%	75%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	39%	80%		

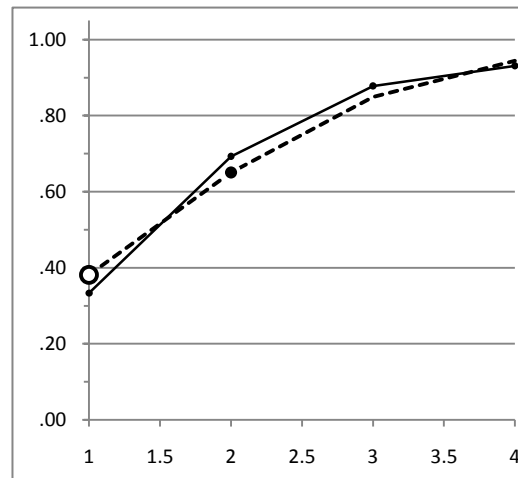
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	329	384	366	451
False Negative	82	167	45	100
True Positive	85	0	55	0
False Positive	55	0	85	0
Sensitivity	.51	.00	.55	.00
specificity	.86	1.00	.81	1.00
Negative Predictive Power	.80	.70	.89	.82
Positive Predictive Power	.61	--	.39	--
Accurate Classification	.75	.70	.76	.82
Kappa	.38	.00	.31	.00

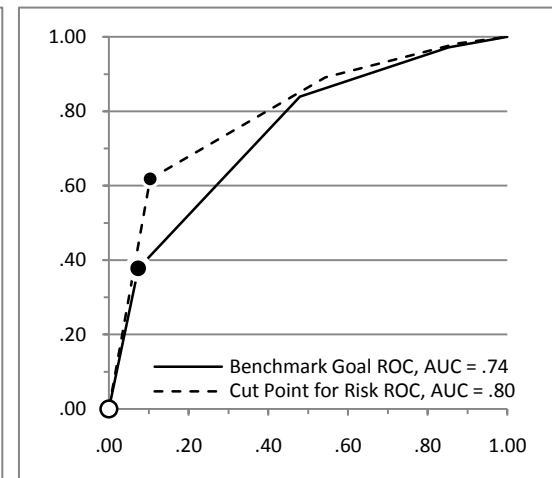
Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	rqr4m	2	1	DORF Retell Quality of Response, Grade 4, Middle of Year
	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

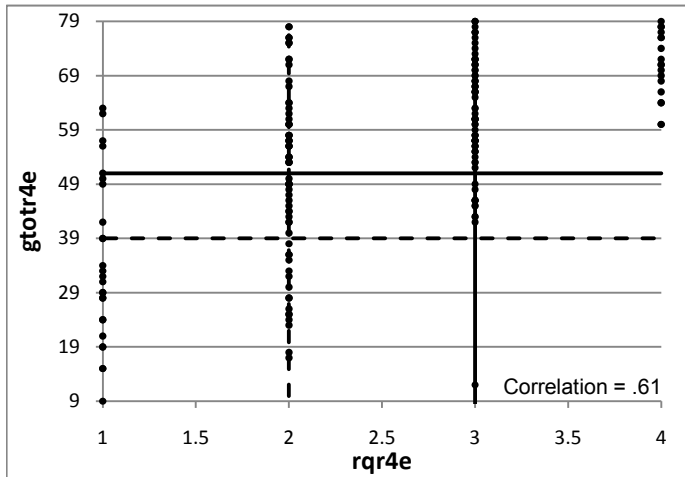
rqr4m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	0	27	340	367	72%
Below Benchmark	0	20	68	88	17%
Well Below Benchmark	0	34	21	55	11%
Marginal Total	0	81	429	510	
Marginal Percent	0%	16%	84%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	33%	79%		

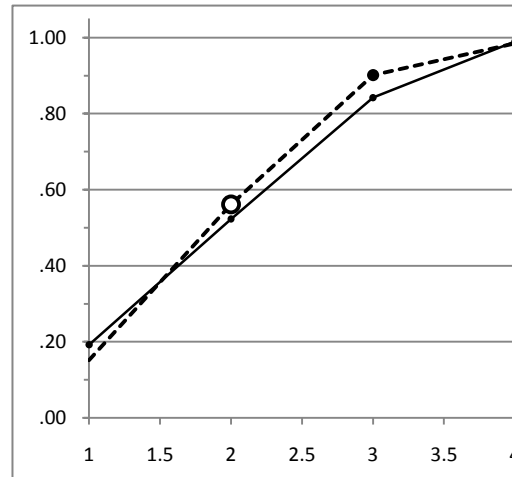
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	340	367	408	455
False Negative	89	143	21	55
True Positive	54	0	34	0
False Positive	27	0	47	0
Sensitivity	.38	.00	.62	.00
specificity	.93	1.00	.90	1.00
Negative Predictive Power	.79	.72	.95	.89
Positive Predictive Power	.67	--	.42	--
Accurate Classification	.77	.72	.87	.89
Kappa	.35	.00	.43	.00

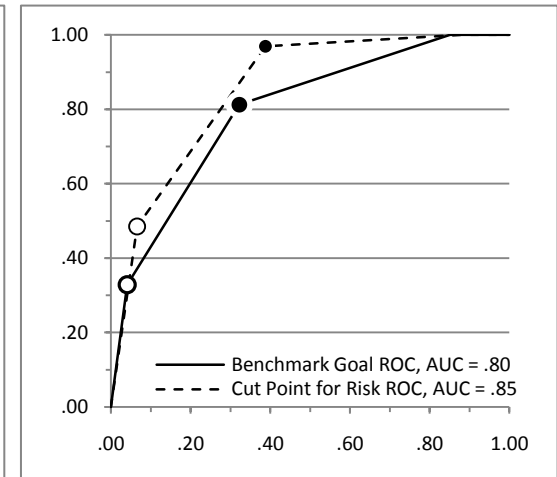
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr4e	3	2	DORF Retell Quality of Response, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

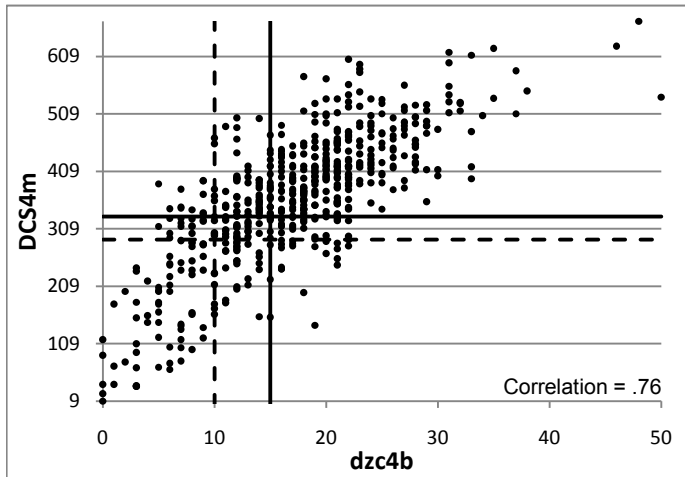
rqr4e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr4e Outcome:					
At or Above Benchmark	5	34	82	121	65%
Below Benchmark	5	15	11	31	17%
Well Below Benchmark	16	16	1	33	18%
Marginal Total	26	65	94	185	
Marginal Percent	14%	35%	51%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	19%	52%	87%		

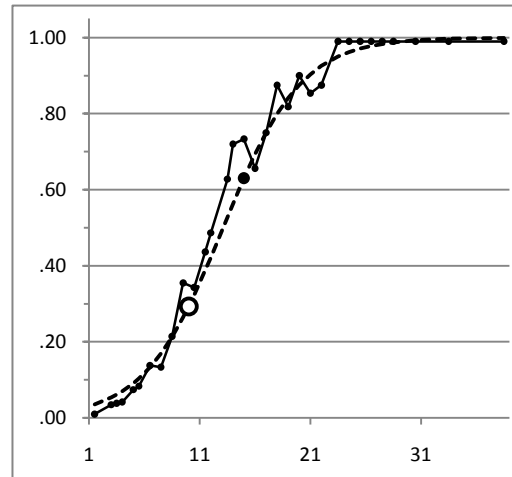
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	82	116	93	142
False Negative	12	43	1	17
True Positive	52	21	32	16
False Positive	39	5	59	10
Sensitivity	.81	.33	.97	.48
specificity	.68	.96	.61	.93
Negative Predictive Power	.87	.73	.99	.89
Positive Predictive Power	.57	.81	.35	.62
Accurate Classification	.72	.74	.68	.85
Kappa	.45	.33	.34	.46

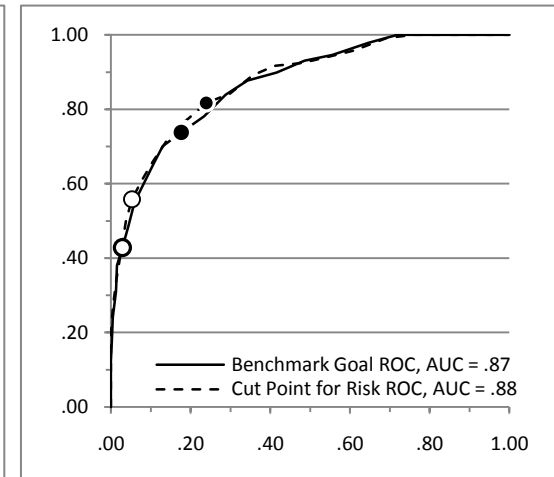
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc4b	15	10	Daze Adjusted Score, Grade 4, Beginning of Year
Outcome	Criterion	DCS4m	330	290	DIBELS Composite Score, Grade 4, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

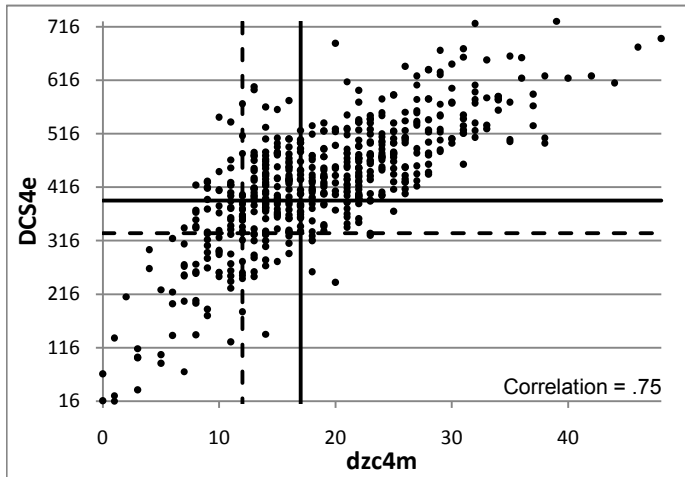
dzc4b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4m Outcome:					
At or Above Benchmark	11	57	317	385	67%
Below Benchmark	13	27	27	67	12%
Well Below Benchmark	67	31	22	120	21%
Marginal Total	91	115	366	572	
Marginal Percent	16%	20%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	12%	50%	87%		

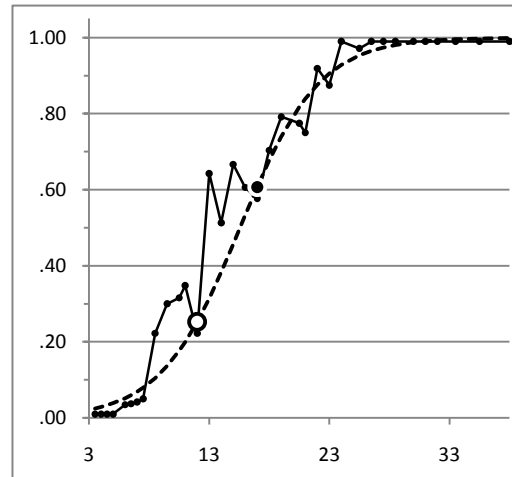
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	317	374	344	428
False Negative	49	107	22	53
True Positive	138	80	98	67
False Positive	68	11	108	24
Sensitivity	.74	.43	.82	.56
specificity	.82	.97	.76	.95
Negative Predictive Power	.87	.78	.94	.89
Positive Predictive Power	.67	.88	.48	.74
Accurate Classification	.80	.79	.77	.87
Kappa	.55	.46	.46	.55

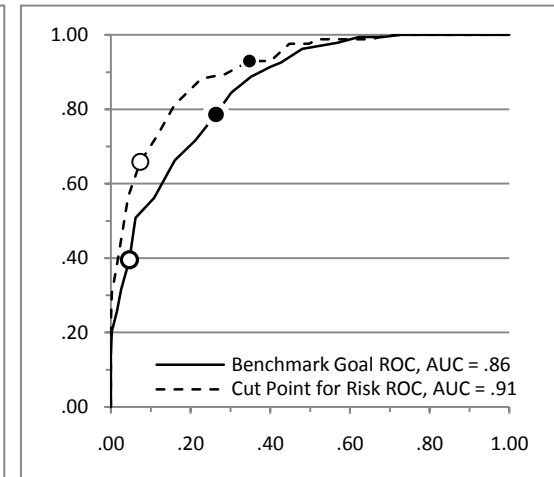
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc4m	17	12	Daze Adjusted Score, Grade 4, Middle of Year
Outcome	Criterion	DCS4e	391	330	DIBELS Composite Score, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

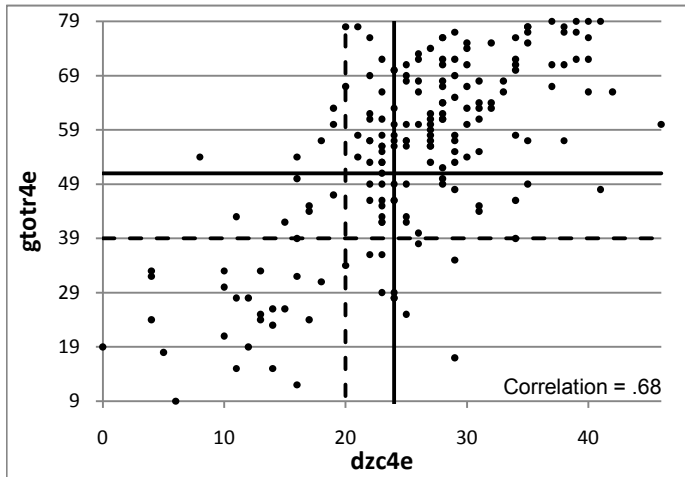
dzc4m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS4e Outcome:					
At or Above Benchmark	18	84	285	387	67%
Below Benchmark	18	50	34	102	18%
Well Below Benchmark	56	23	6	85	15%
Marginal Total	92	157	325	574	
Marginal Percent	16%	27%	57%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	54%	88%		

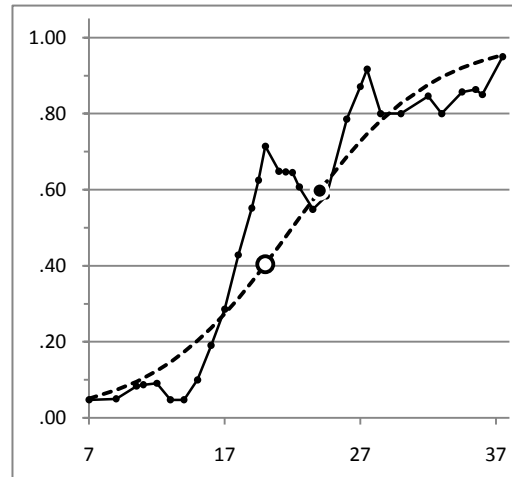
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	285	369	319	453
False Negative	40	113	6	29
True Positive	147	74	79	56
False Positive	102	18	170	36
Sensitivity	.79	.40	.93	.66
specificity	.74	.95	.65	.93
Negative Predictive Power	.88	.77	.98	.94
Positive Predictive Power	.59	.80	.32	.61
Accurate Classification	.75	.77	.69	.89
Kappa	.48	.40	.32	.57

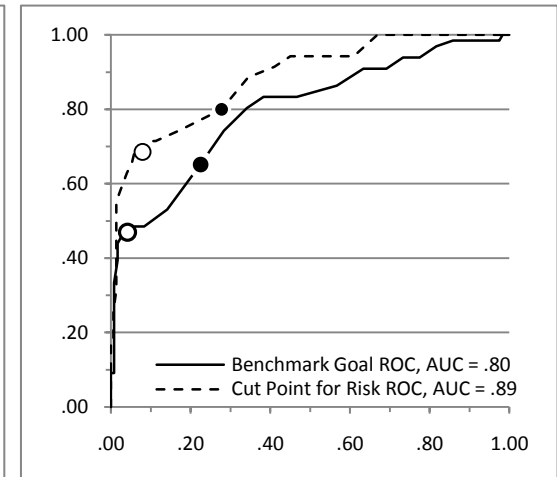
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc4e	24	20	Daze Adjusted Score, Grade 4, End of Year
Outcome	Criterion	gtotr4e	51	39	GRADE Total Test, Grade 4, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

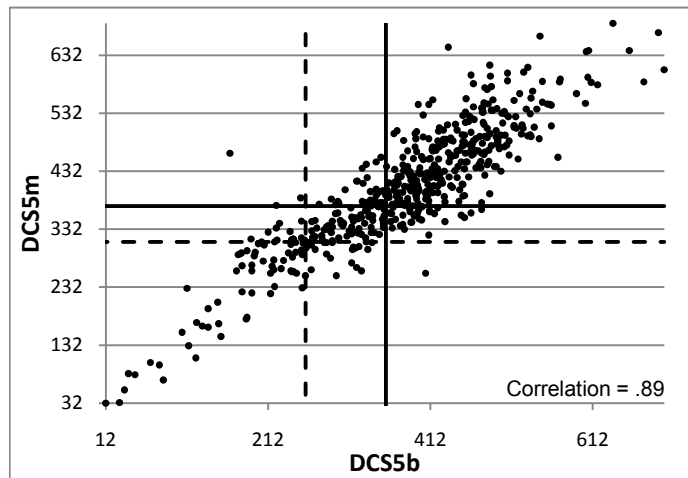
dzc4e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr4e Outcome:					
At or Above Benchmark	5	22	93	120	65%
Below Benchmark	7	8	16	31	17%
Well Below Benchmark	24	4	7	35	19%
Marginal Total	36	34	116	186	
Marginal Percent	19%	18%	62%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	65%	80%		

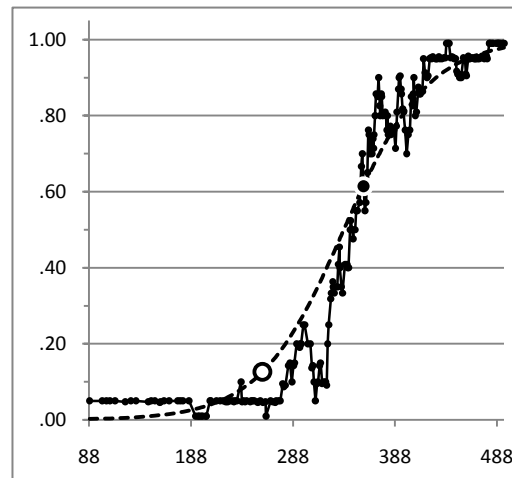
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	93	115	109	139
False Negative	23	35	7	11
True Positive	43	31	28	24
False Positive	27	5	42	12
Sensitivity	.65	.47	.80	.69
specificity	.78	.96	.72	.92
Negative Predictive Power	.80	.77	.94	.93
Positive Predictive Power	.61	.86	.40	.67
Accurate Classification	.73	.78	.74	.88
Kappa	.42	.48	.38	.60

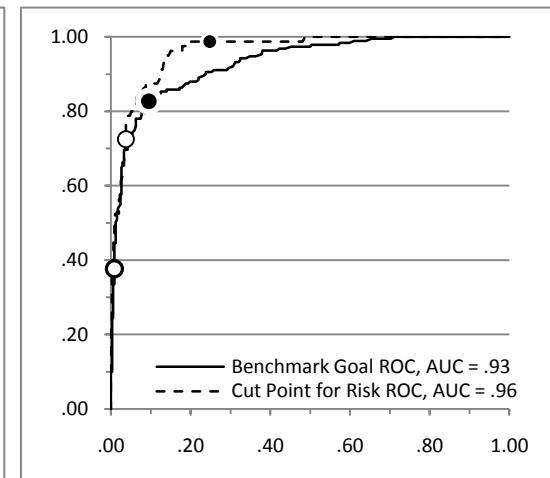
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS5b	357	258	DIBELS Composite Score, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

DCS5b Screening Decision:

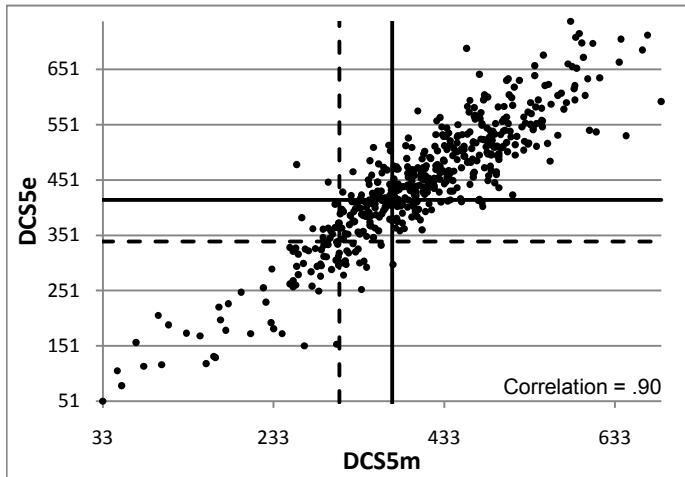
DCS5m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	3	29	304	336	64%
Below Benchmark	14	65	32	111	21%
Well Below Benchmark	58	21	1	80	15%
Marginal Total	75	115	337	527	
Marginal Percent	14%	22%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	4%	25%	90%		

At or Above Benchmark outcome

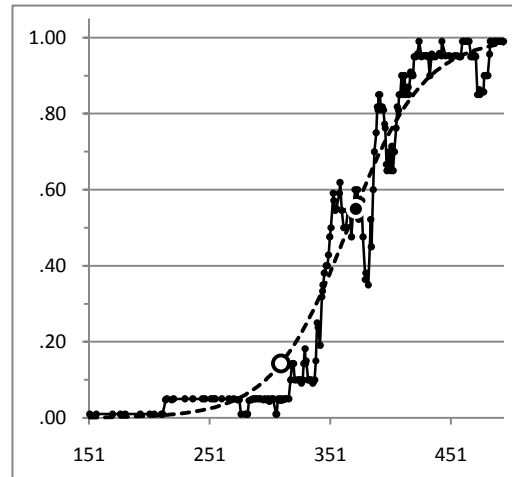
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	304	333	336	430
False Negative	33	119	1	22
True Positive	158	72	79	58
False Positive	32	3	111	17
Sensitivity	.83	.38	.99	.73
specificity	.90	.99	.75	.96
Negative Predictive Power	.90	.74	1.00	.95
Positive Predictive Power	.83	.96	.42	.77
Accurate Classification	.88	.77	.79	.93
Kappa	.73	.42	.47	.71

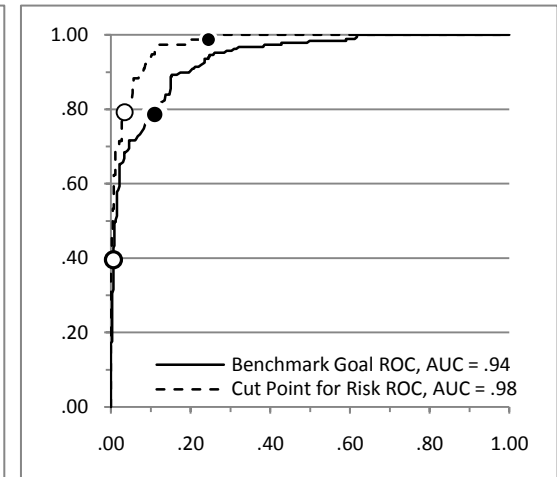
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

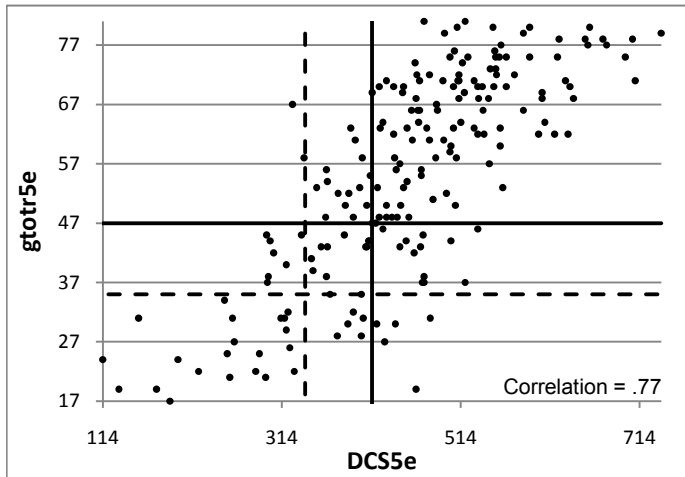
DCS5m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	2	34	291	327	64%
Below Benchmark	13	58	39	110	21%
Well Below Benchmark	61	15	1	77	15%
Marginal Total	76	107	331	514	
Marginal Percent	15%	21%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	3%	32%	88%		

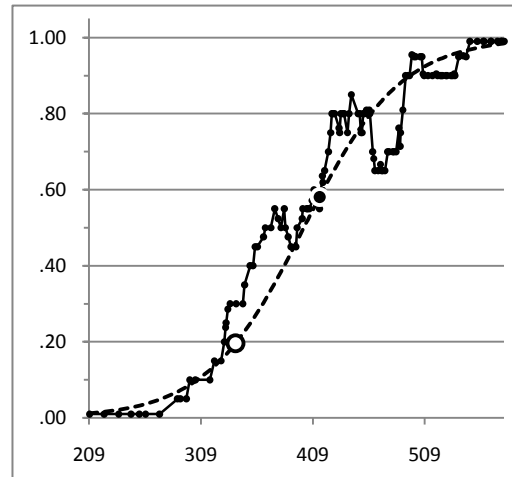
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	291	325	330	422
False Negative	40	113	1	16
True Positive	147	74	76	61
False Positive	36	2	107	15
Sensitivity	.79	.40	.99	.79
specificity	.89	.99	.76	.97
Negative Predictive Power	.88	.74	1.00	.96
Positive Predictive Power	.80	.97	.42	.80
Accurate Classification	.85	.78	.79	.94
Kappa	.68	.45	.47	.76

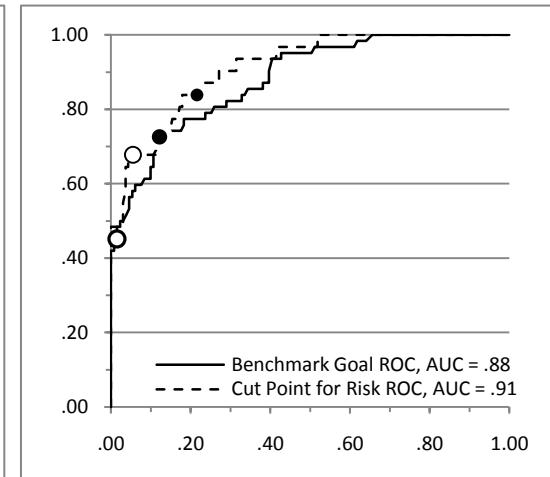
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

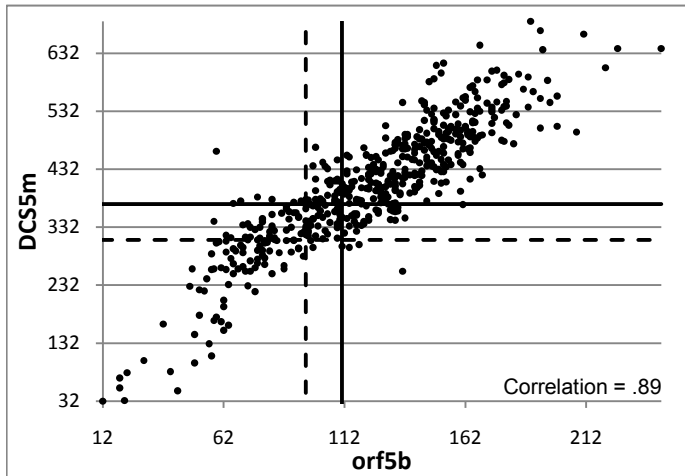
DCS5e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	2	14	115	131	68%
Below Benchmark	7	12	12	31	16%
Well Below Benchmark	21	5	5	31	16%
Marginal Total	30	31	132	193	
Marginal Percent	16%	16%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	7%	45%	87%		

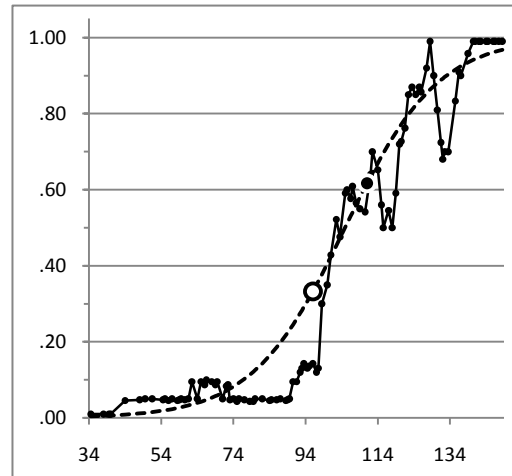
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	115	129	127	153
False Negative	17	34	5	10
True Positive	45	28	26	21
False Positive	16	2	35	9
Sensitivity	.73	.45	.84	.68
specificity	.88	.98	.78	.94
Negative Predictive Power	.87	.79	.96	.94
Positive Predictive Power	.74	.93	.43	.70
Accurate Classification	.83	.81	.79	.90
Kappa	.61	.50	.45	.63

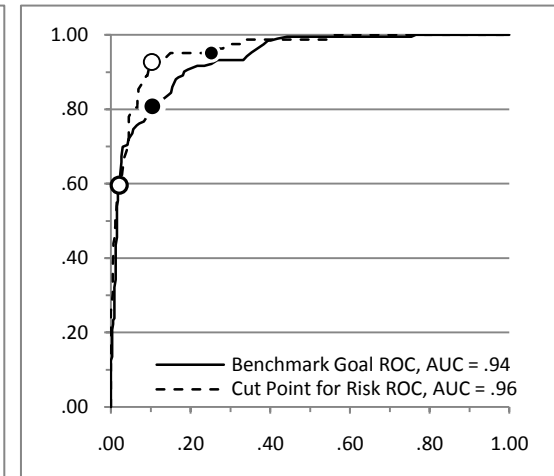
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5b	111	96	DORF Words Correct, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

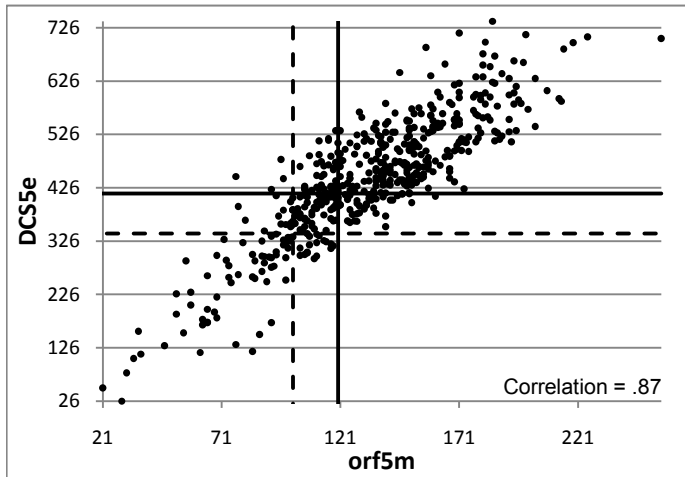
orf5b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5m Outcome:					
At or Above Benchmark	7	28	302	337	64%
Below Benchmark	39	39	33	111	21%
Well Below Benchmark	76	2	4	82	15%
Marginal Total	122	69	339	530	
Marginal Percent	23%	13%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	41%	89%		

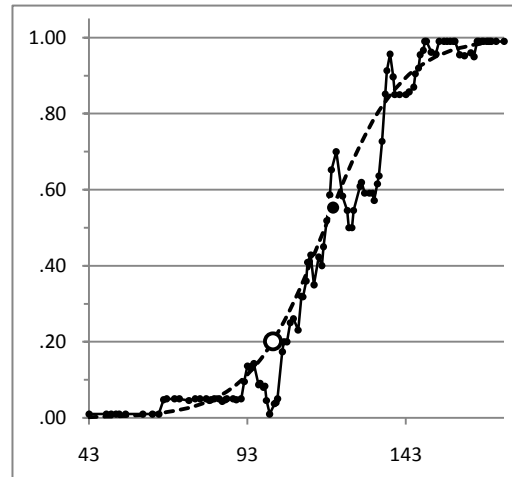
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	302	330	335	402
False Negative	37	78	4	6
True Positive	156	115	78	76
False Positive	35	7	113	46
Sensitivity	.81	.60	.95	.93
specificity	.90	.98	.75	.90
Negative Predictive Power	.89	.81	.99	.99
Positive Predictive Power	.82	.94	.41	.62
Accurate Classification	.86	.84	.78	.90
Kappa	.71	.62	.45	.69

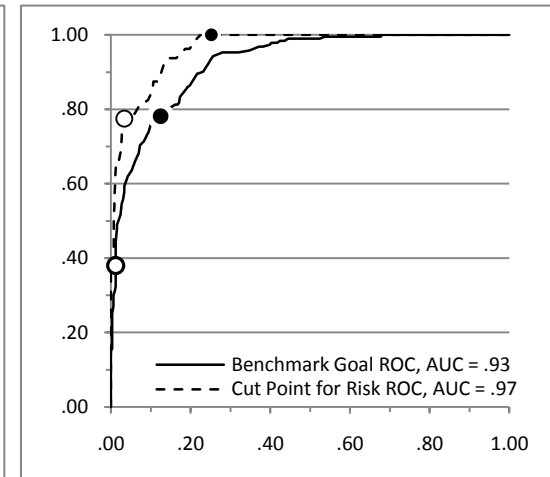
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5m	120	101	DORF Words Correct, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

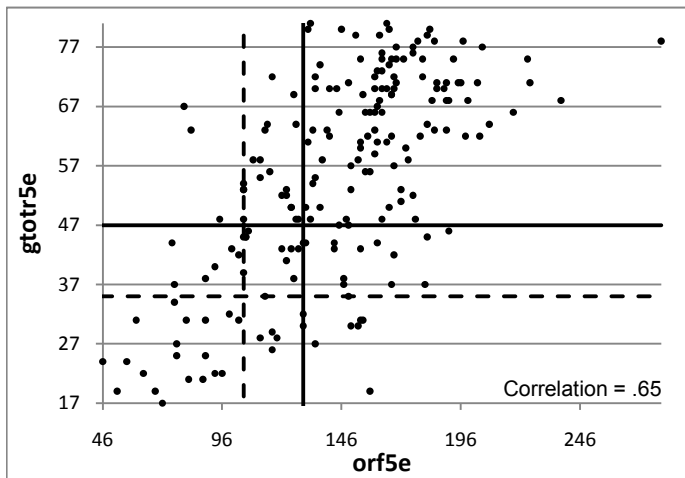
orf5m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	4	37	287	328	63%
Below Benchmark	11	59	42	112	22%
Well Below Benchmark	62	18	0	80	15%
Marginal Total	77	114	329	520	
Marginal Percent	15%	22%	63%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	5%	32%	87%		

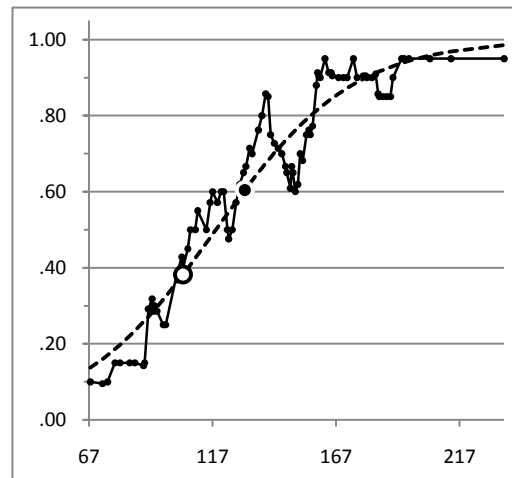
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	287	324	329	425
False Negative	42	119	0	18
True Positive	150	73	80	62
False Positive	41	4	111	15
Sensitivity	.78	.38	1.00	.78
specificity	.88	.99	.75	.97
Negative Predictive Power	.87	.73	1.00	.96
Positive Predictive Power	.79	.95	.42	.81
Accurate Classification	.84	.76	.79	.94
Kappa	.66	.42	.48	.75

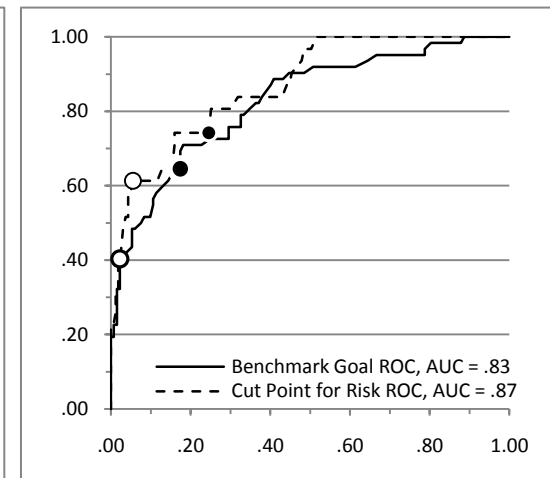
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5e	130	105	DORF Words Correct, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

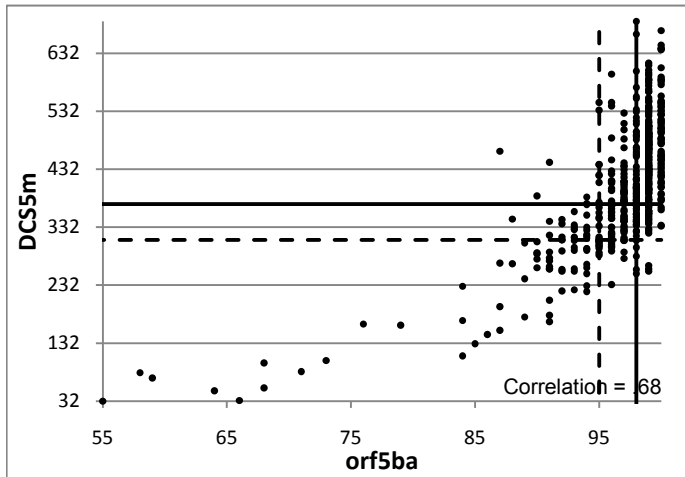
orf5e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	3	20	109	132	68%
Below Benchmark	6	11	14	31	16%
Well Below Benchmark	19	4	8	31	16%
Marginal Total	28	35	131	194	
Marginal Percent	14%	18%	68%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	11%	57%	83%		

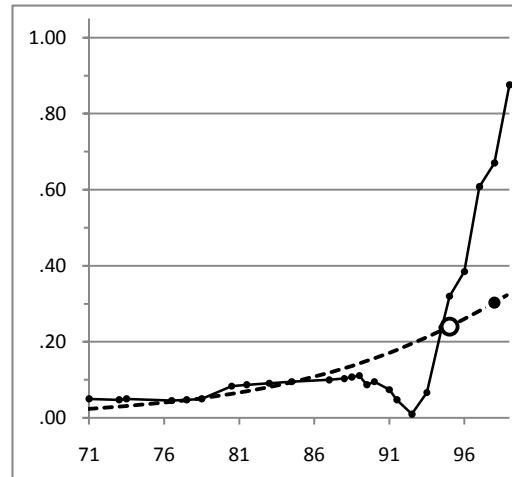
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	109	129	123	154
False Negative	22	37	8	12
True Positive	40	25	23	19
False Positive	23	3	40	9
Sensitivity	.65	.40	.74	.61
specificity	.83	.98	.75	.94
Negative Predictive Power	.83	.78	.94	.93
Positive Predictive Power	.63	.89	.37	.68
Accurate Classification	.77	.79	.75	.89
Kappa	.47	.45	.35	.58

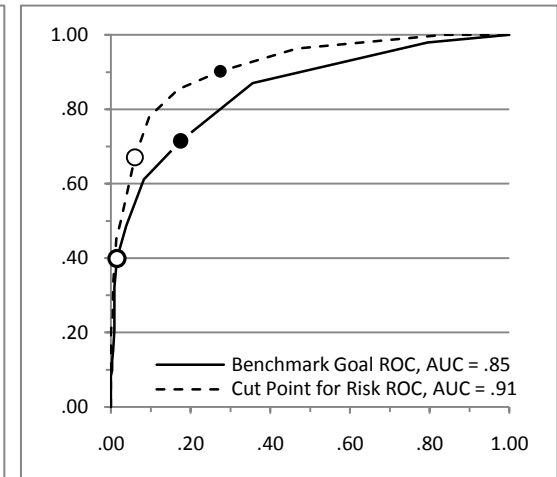
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5ba	98	95	DORF Accuracy, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf5ba Screening Decision:

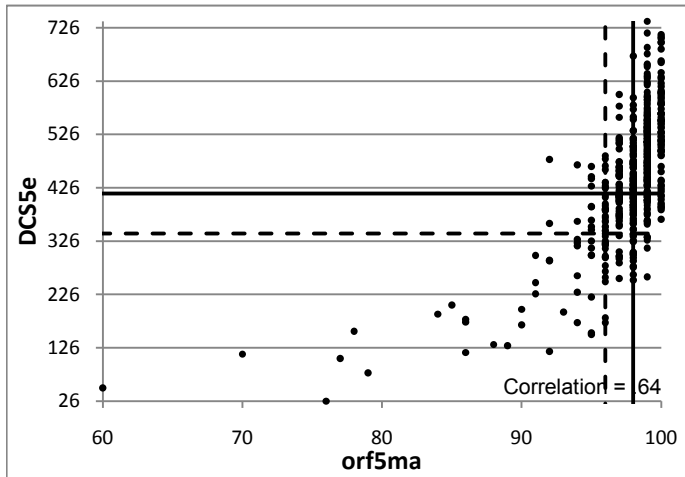
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5m Outcome:					
At or Above Benchmark	5	54	278	337	64%
Below Benchmark	22	42	47	111	21%
Well Below Benchmark	55	19	8	82	15%
Marginal Total	82	115	333	530	
Marginal Percent	15%	22%	63%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	47%	83%		

At or Above Benchmark outcome

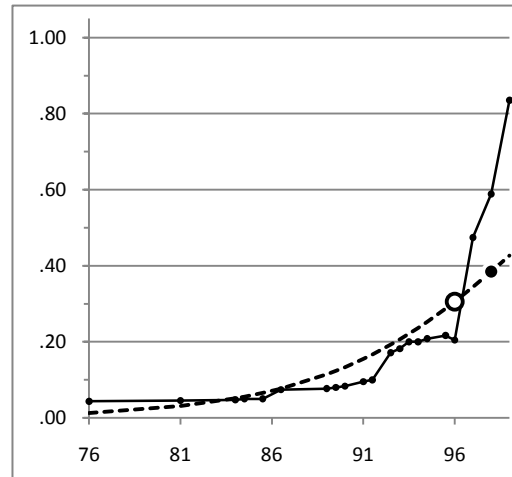
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	278	332	325	421
False Negative	55	116	8	27
True Positive	138	77	74	55
False Positive	59	5	123	27
Sensitivity	.72	.40	.90	.67
specificity	.82	.99	.73	.94
Negative Predictive Power	.83	.74	.98	.94
Positive Predictive Power	.70	.94	.38	.67
Accurate Classification	.78	.77	.75	.90
Kappa	.54	.44	.40	.61

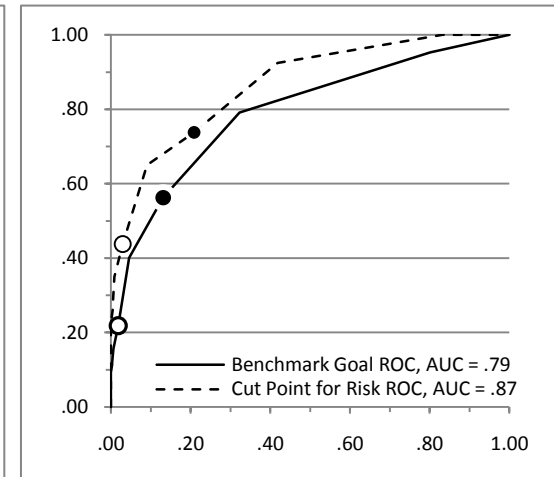
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5ma	98	96	DORF Accuracy, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

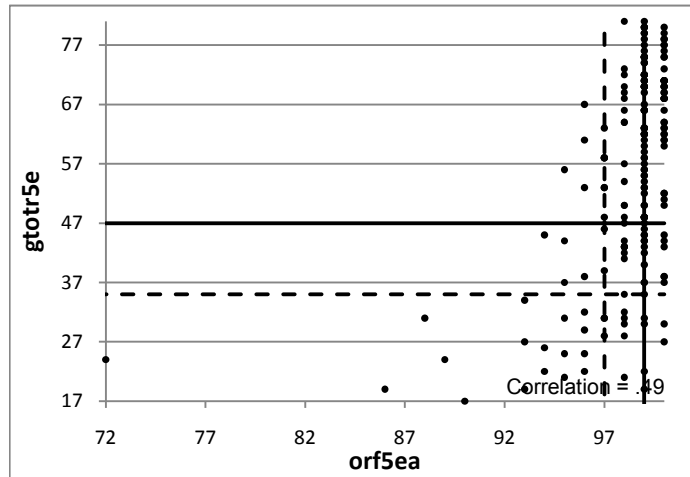
orf5ma Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	6	37	285	328	63%
Below Benchmark	7	42	63	112	22%
Well Below Benchmark	35	24	21	80	15%
Marginal Total	48	103	369	520	
Marginal Percent	9%	20%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	13%	36%	77%		

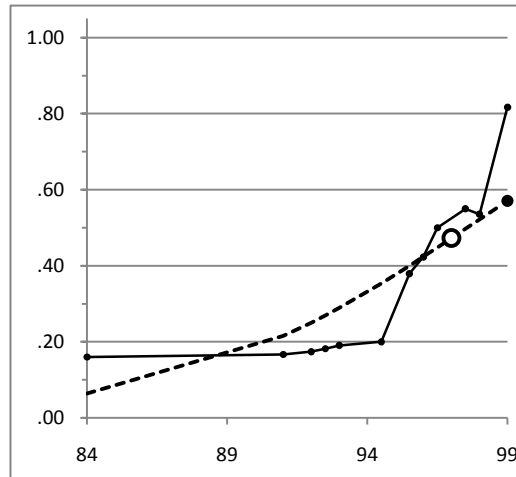
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	285	322	348	427
False Negative	84	150	21	45
True Positive	108	42	59	35
False Positive	43	6	92	13
Sensitivity	.56	.22	.74	.44
specificity	.87	.98	.79	.97
Negative Predictive Power	.77	.68	.94	.90
Positive Predictive Power	.72	.88	.39	.73
Accurate Classification	.76	.70	.78	.89
Kappa	.45	.24	.39	.49

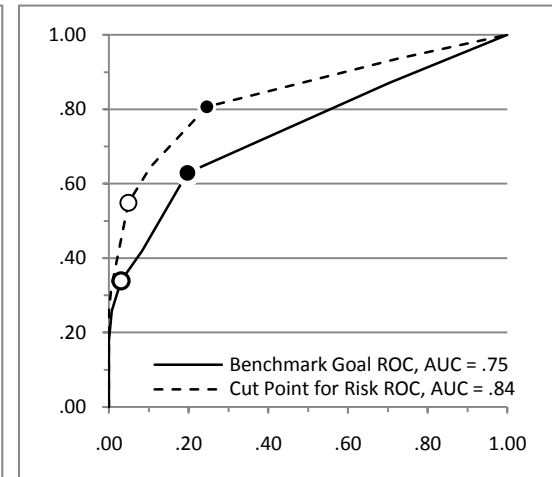
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf5ea	99	97	DORF Accuracy, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf5ea Screening Decision:

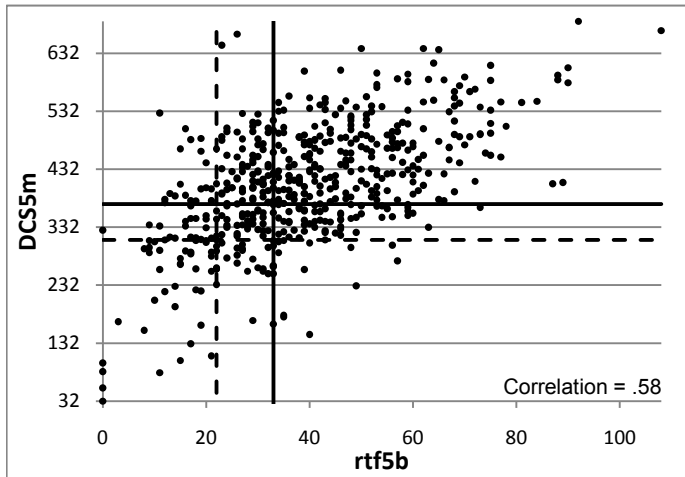
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	4	22	106	132	68%
Below Benchmark	4	10	17	31	16%
Well Below Benchmark	17	8	6	31	16%
Marginal Total	25	40	129	194	
Marginal Percent	13%	21%	66%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	16%	55%	82%		

At or Above Benchmark outcome

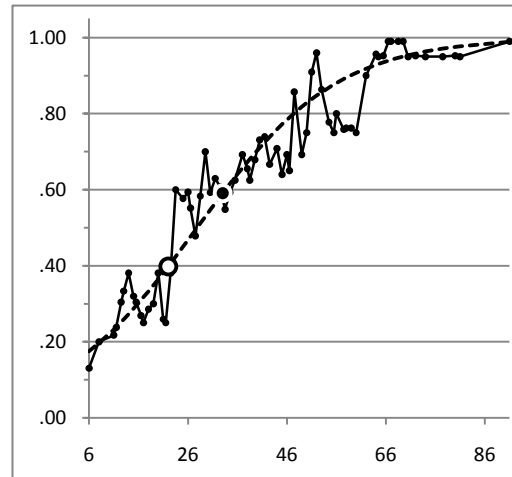
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	106	128	123	155
False Negative	23	41	6	14
True Positive	39	21	25	17
False Positive	26	4	40	8
Sensitivity	.63	.34	.81	.55
specificity	.80	.97	.75	.95
Negative Predictive Power	.82	.76	.95	.92
Positive Predictive Power	.60	.84	.38	.68
Accurate Classification	.75	.77	.76	.89
Kappa	.43	.37	.39	.54

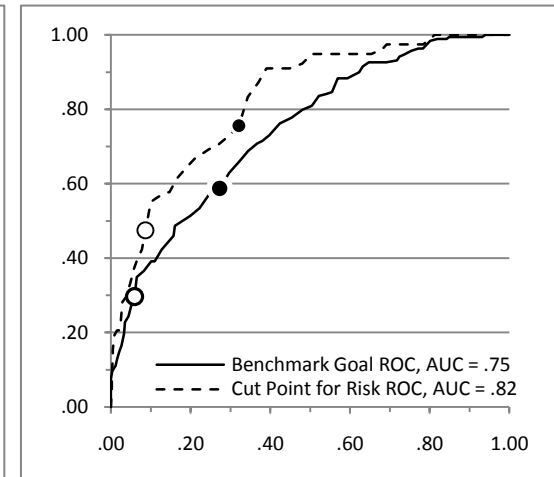
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf5b	33	22	DORF Retell, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

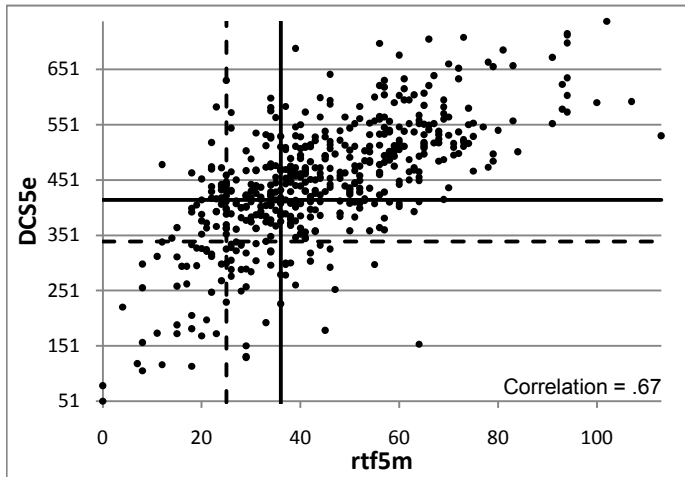
rtf5b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5m Outcome:					
At or Above Benchmark	20	72	245	337	64%
Below Benchmark	19	33	59	111	21%
Well Below Benchmark	37	22	19	78	15%
Marginal Total	76	127	323	526	
Marginal Percent	14%	24%	61%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	26%	57%	76%		

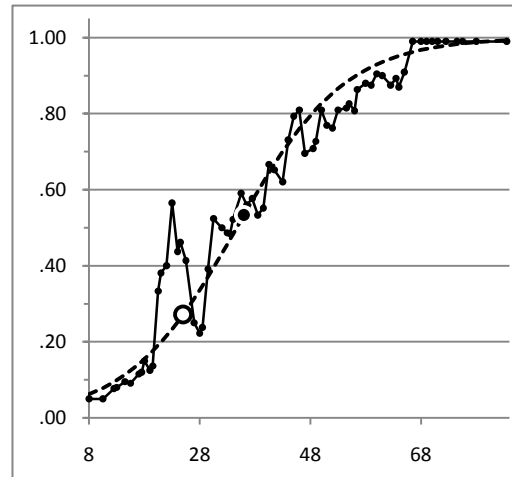
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	245	317	304	409
False Negative	78	133	19	41
True Positive	111	56	59	37
False Positive	92	20	144	39
Sensitivity	.59	.30	.76	.47
specificity	.73	.94	.68	.91
Negative Predictive Power	.76	.70	.94	.91
Positive Predictive Power	.55	.74	.29	.49
Accurate Classification	.68	.71	.69	.85
Kappa	.31	.27	.26	.39

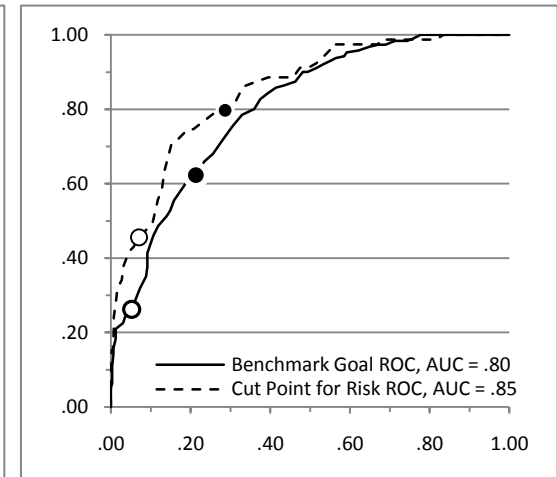
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf5m	36	25	DORF Retell, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

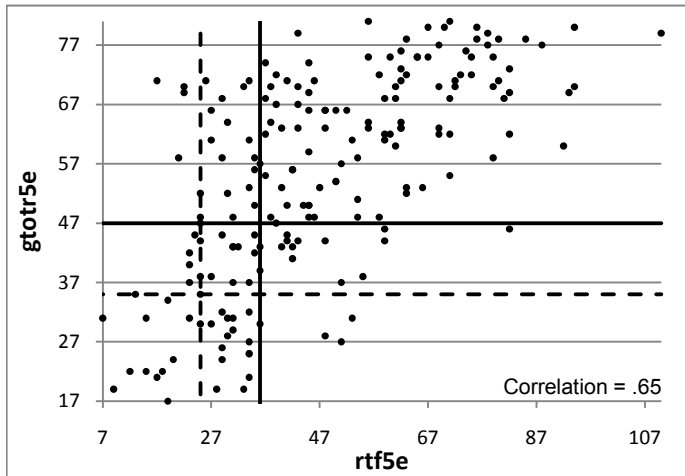
rtf5m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	17	53	258	328	63%
Below Benchmark	14	42	56	112	22%
Well Below Benchmark	36	27	16	79	15%
Marginal Total	67	122	330	519	
Marginal Percent	13%	24%	64%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	25%	43%	78%		

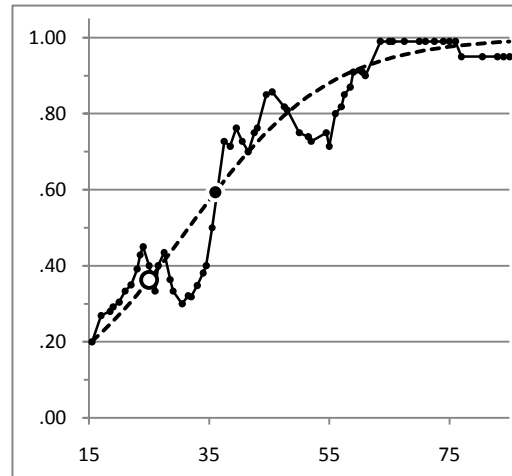
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	258	311	314	409
False Negative	72	141	16	43
True Positive	119	50	63	36
False Positive	70	17	126	31
Sensitivity	.62	.26	.80	.46
specificity	.79	.95	.71	.93
Negative Predictive Power	.78	.69	.95	.90
Positive Predictive Power	.63	.75	.33	.54
Accurate Classification	.73	.70	.73	.86
Kappa	.41	.24	.33	.41

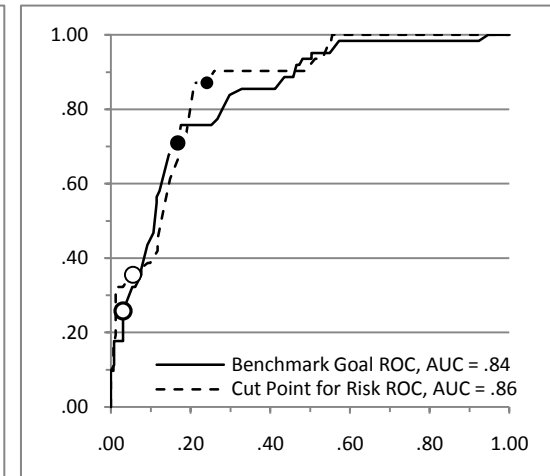
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf5e	36	25	DORF Retell, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rtf5e Screening Decision:

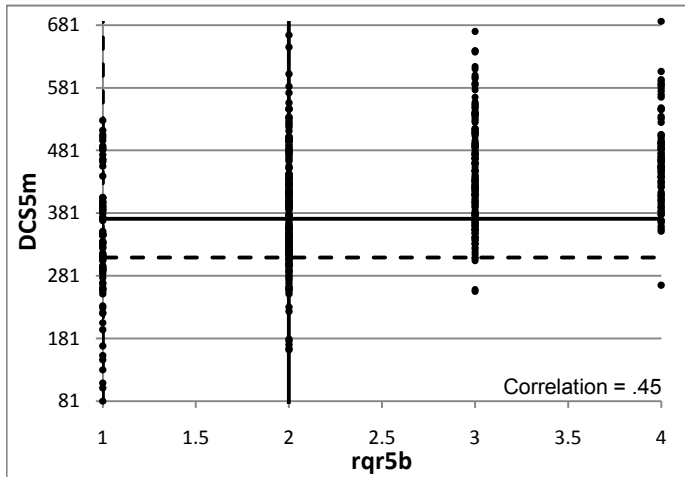
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	4	18	109	131	68%
Below Benchmark	5	12	14	31	16%
Well Below Benchmark	11	16	4	31	16%
Marginal Total	20	46	127	193	
Marginal Percent	10%	24%	66%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	39%	86%		

At or Above Benchmark outcome

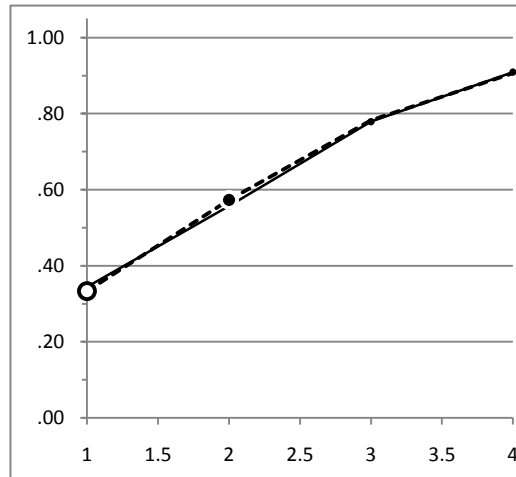
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	109	127	123	153
False Negative	18	46	4	20
True Positive	44	16	27	11
False Positive	22	4	39	9
Sensitivity	.71	.26	.87	.35
specificity	.83	.97	.76	.94
Negative Predictive Power	.86	.73	.97	.88
Positive Predictive Power	.67	.80	.41	.55
Accurate Classification	.79	.74	.78	.85
Kappa	.53	.28	.43	.35

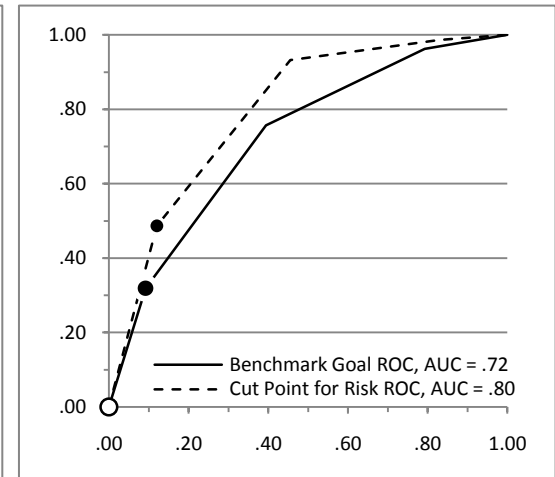
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr5b	2	1	DORF Retell Quality of Response, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

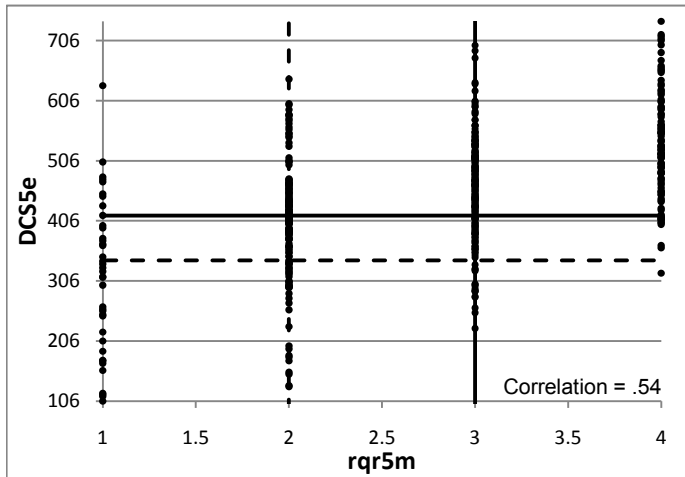
rqr5b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5m Outcome:					
At or Above Benchmark	0	31	306	337	65%
Below Benchmark	0	23	88	111	21%
Well Below Benchmark	0	36	38	74	14%
Marginal Total	0	90	432	522	
Marginal Percent	0%	17%	83%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	34%	71%		

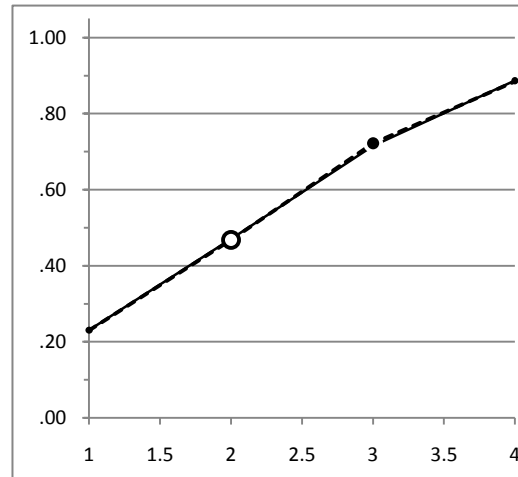
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	306	337	394	448
False Negative	126	185	38	74
True Positive	59	0	36	0
False Positive	31	0	54	0
Sensitivity	.32	.00	.49	.00
specificity	.91	1.00	.88	1.00
Negative Predictive Power	.71	.65	.91	.86
Positive Predictive Power	.66	--	.40	--
Accurate Classification	.70	.65	.82	.86
Kappa	.26	.00	.34	.00

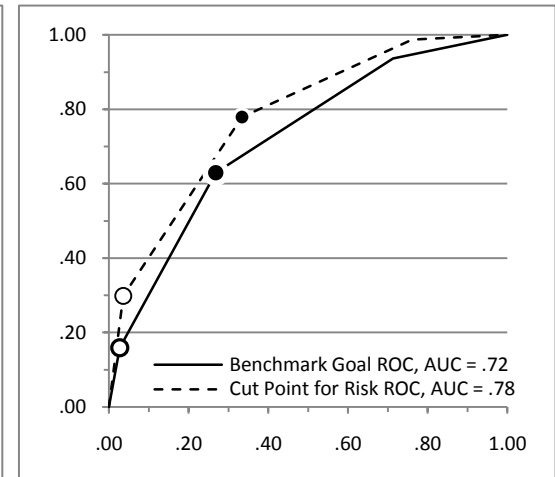
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr5m	3	2	DORF Retell Quality of Response, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

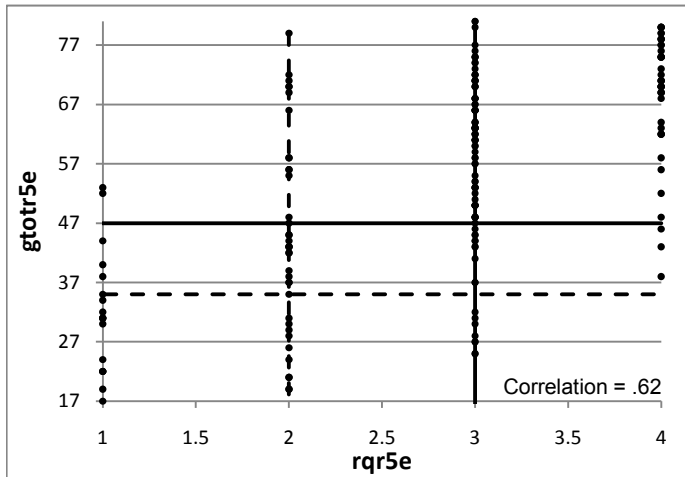
rqr5m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	9	79	240	328	63%
Below Benchmark	7	52	53	112	22%
Well Below Benchmark	23	37	17	77	15%
Marginal Total	39	168	310	517	
Marginal Percent	8%	32%	60%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	23%	47%	77%		

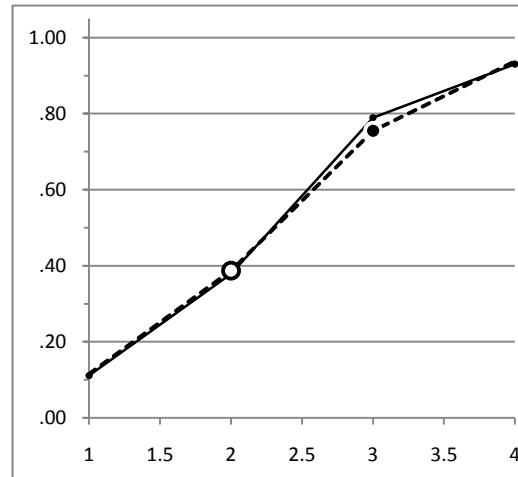
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	240	319	293	424
False Negative	70	159	17	54
True Positive	119	30	60	23
False Positive	88	9	147	16
Sensitivity	.63	.16	.78	.30
specificity	.73	.97	.67	.96
Negative Predictive Power	.77	.67	.95	.89
Positive Predictive Power	.57	.77	.29	.59
Accurate Classification	.69	.68	.68	.86
Kappa	.35	.16	.26	.33

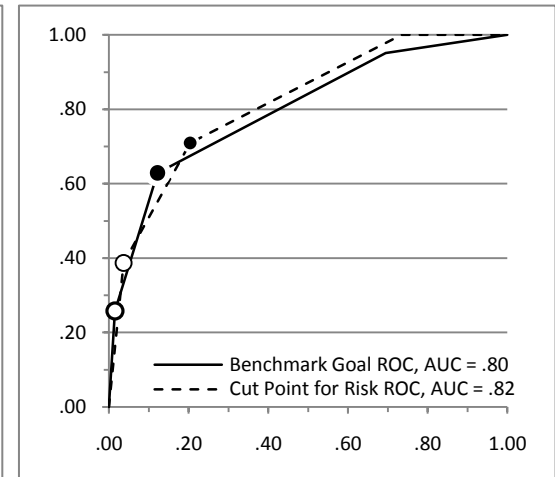
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr5e	3	2	DORF Retell Quality of Response, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rqr5e Screening Decision:

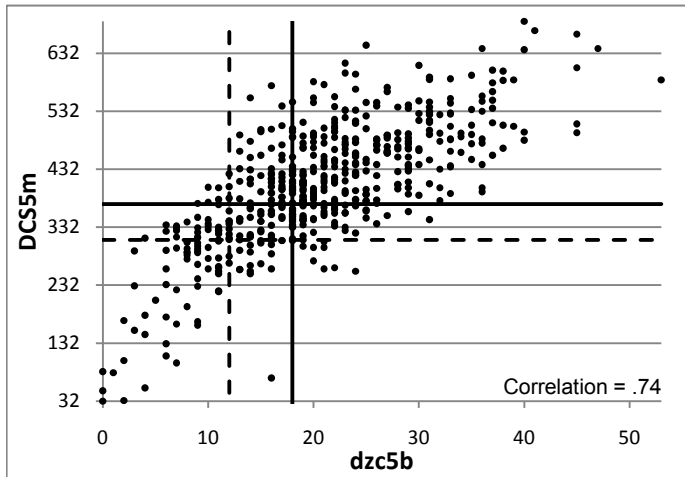
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	2	14	115	131	68%
Below Benchmark	4	13	14	31	16%
Well Below Benchmark	12	10	9	31	16%
Marginal Total	18	37	138	193	
Marginal Percent	9%	19%	72%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	11%	38%	83%		

At or Above Benchmark outcome

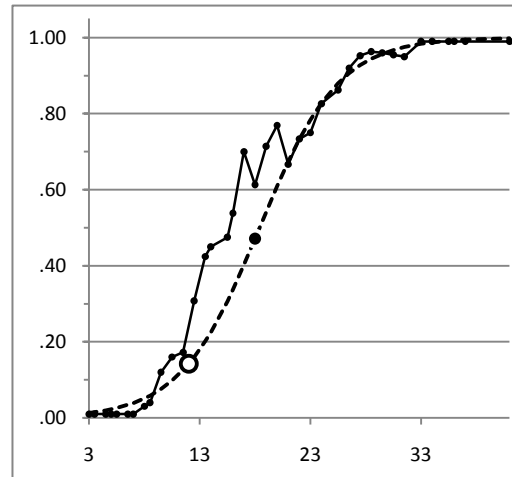
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	115	129	129	156
False Negative	23	46	9	19
True Positive	39	16	22	12
False Positive	16	2	33	6
Sensitivity	.63	.26	.71	.39
specificity	.88	.98	.80	.96
Negative Predictive Power	.83	.74	.93	.89
Positive Predictive Power	.71	.89	.40	.67
Accurate Classification	.80	.75	.78	.87
Kappa	.52	.30	.39	.42

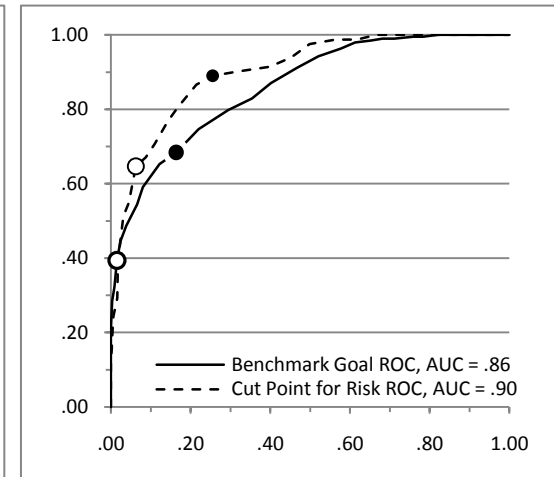
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc5b	18	12	Daze Adjusted Score, Grade 5, Beginning of Year
Outcome	Criterion	DCS5m	372	310	DIBELS Composite Score, Grade 5, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

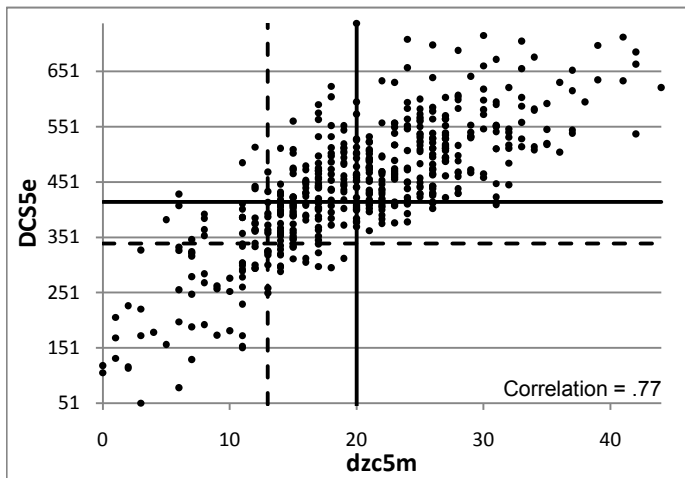
dzc5b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5m Outcome:					
At or Above Benchmark	5	50	281	336	64%
Below Benchmark	23	36	52	111	21%
Well Below Benchmark	53	20	9	82	16%
Marginal Total	81	106	342	529	
Marginal Percent	15%	20%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	47%	82%		

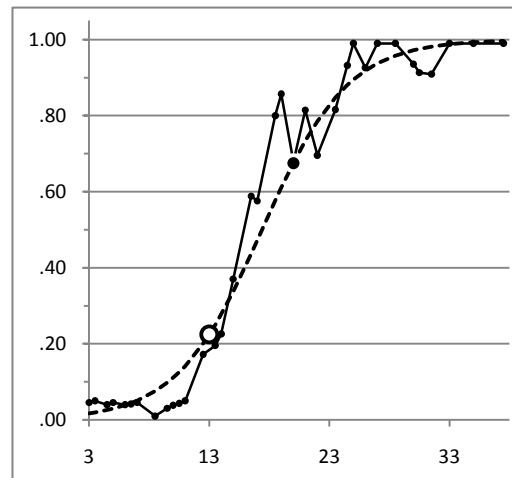
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	281	331	333	419
False Negative	61	117	9	29
True Positive	132	76	73	53
False Positive	55	5	114	28
Sensitivity	.68	.39	.89	.65
specificity	.84	.99	.74	.94
Negative Predictive Power	.82	.74	.97	.94
Positive Predictive Power	.71	.94	.39	.65
Accurate Classification	.78	.77	.77	.89
Kappa	.52	.43	.42	.59

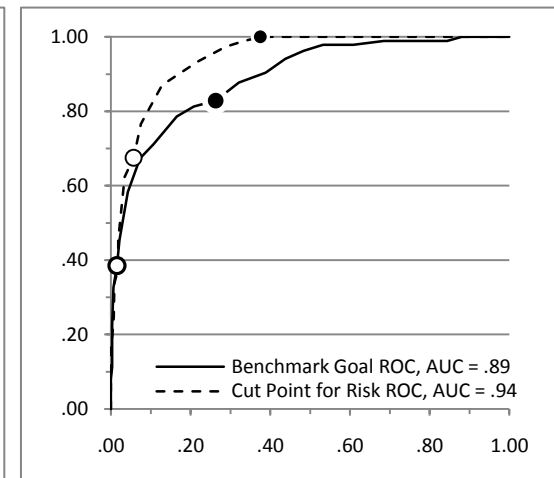
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc5m	20	13	Daze Adjusted Score, Grade 5, Middle of Year
Outcome	Criterion	DCS5e	415	340	DIBELS Composite Score, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

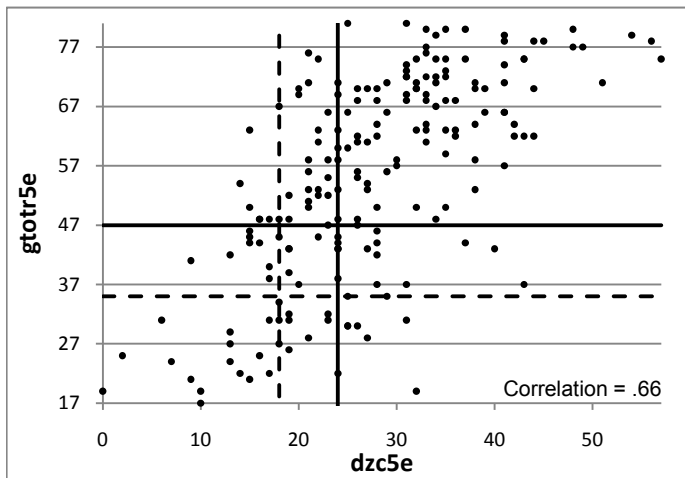
dzc5m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS5e Outcome:					
At or Above Benchmark	5	81	241	327	64%
Below Benchmark	20	58	32	110	21%
Well Below Benchmark	52	25	0	77	15%
Marginal Total	77	164	273	514	
Marginal Percent	15%	32%	53%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	6%	49%	88%		

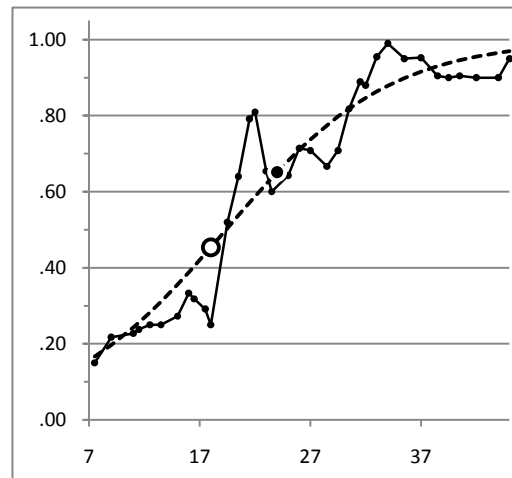
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	241	322	273	412
False Negative	32	115	0	25
True Positive	155	72	77	52
False Positive	86	5	164	25
Sensitivity	.83	.39	1.00	.68
specificity	.74	.98	.62	.94
Negative Predictive Power	.88	.74	1.00	.94
Positive Predictive Power	.64	.94	.32	.68
Accurate Classification	.77	.77	.68	.90
Kappa	.53	.42	.33	.62

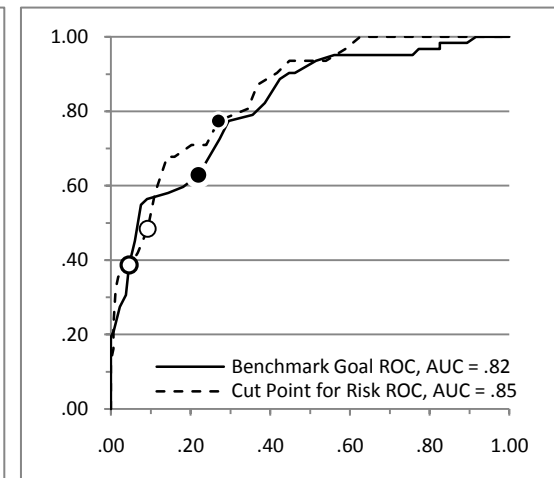
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc5e	24	18	Daze Adjusted Score, Grade 5, End of Year
Outcome	Criterion	gtotr5e	47	35	GRADE Total Test, Grade 5, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

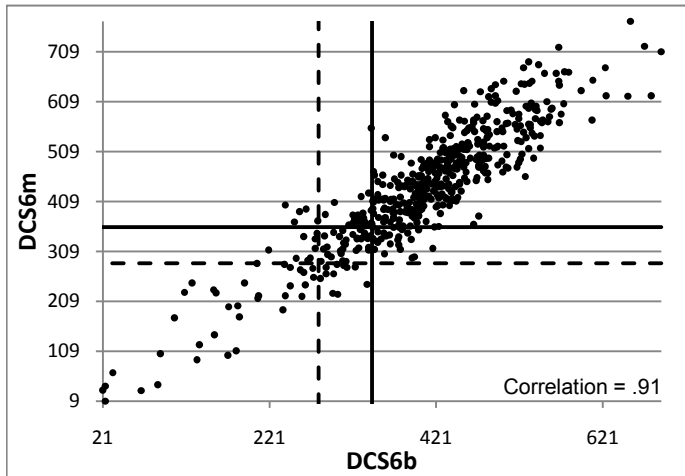
dzc5e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr5e Outcome:					
At or Above Benchmark	6	23	103	132	68%
Below Benchmark	9	6	16	31	16%
Well Below Benchmark	15	9	7	31	16%
Marginal Total	30	38	126	194	
Marginal Percent	15%	20%	65%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	61%	82%		

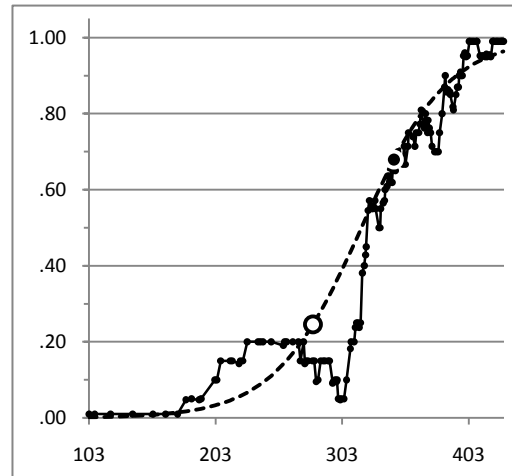
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	103	126	119	148
False Negative	23	38	7	16
True Positive	39	24	24	15
False Positive	29	6	44	15
Sensitivity	.63	.39	.77	.48
specificity	.78	.95	.73	.91
Negative Predictive Power	.82	.77	.94	.90
Positive Predictive Power	.57	.80	.35	.50
Accurate Classification	.73	.77	.74	.84
Kappa	.40	.40	.34	.40

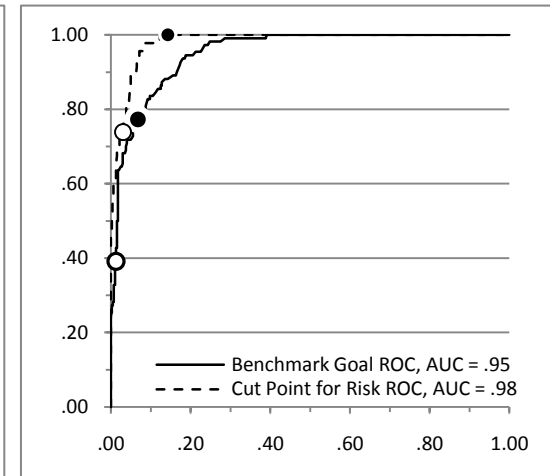
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS6b	344	280	DIBELS Composite Score, Grade 6, Beginning of Year
Outcome	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

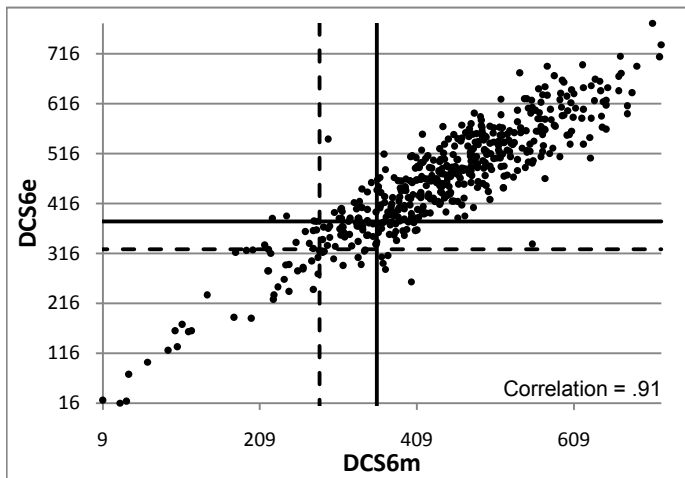
DCS6b Screening Decision:

DCS6m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	5	22	372	399	78%
Below Benchmark	9	30	25	64	13%
Well Below Benchmark	34	12	0	46	9%
Marginal Total	48	64	397	509	
Marginal Percent	9%	13%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	10%	34%	94%		

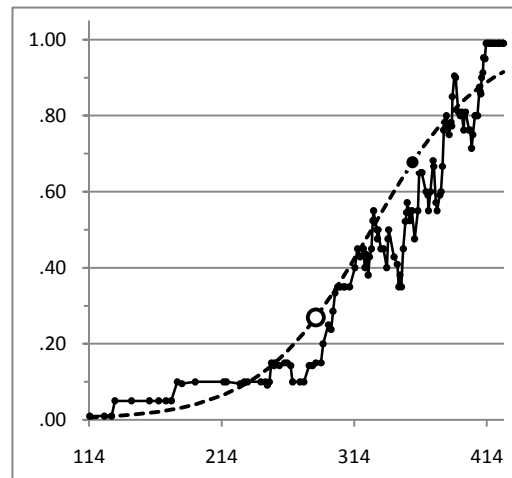
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	372	394	397	449
False Negative	25	67	0	12
True Positive	85	43	46	34
False Positive	27	5	66	14
Sensitivity	.77	.39	1.00	.74
specificity	.93	.99	.86	.97
Negative Predictive Power	.94	.85	1.00	.97
Positive Predictive Power	.76	.90	.41	.71
Accurate Classification	.90	.86	.87	.95
Kappa	.70	.48	.52	.70

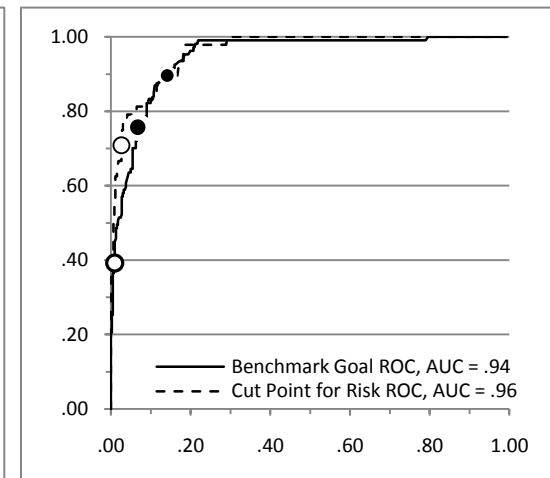
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year
Outcome	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

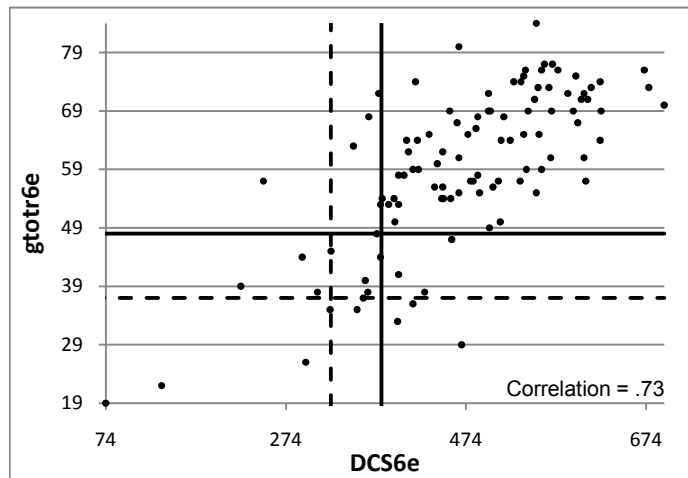
DCS6m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6e Outcome:					
At or Above Benchmark	4	23	374	401	79%
Below Benchmark	8	30	21	59	12%
Well Below Benchmark	34	9	5	48	9%
Marginal Total	46	62	400	508	
Marginal Percent	9%	12%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	9%	37%	94%		

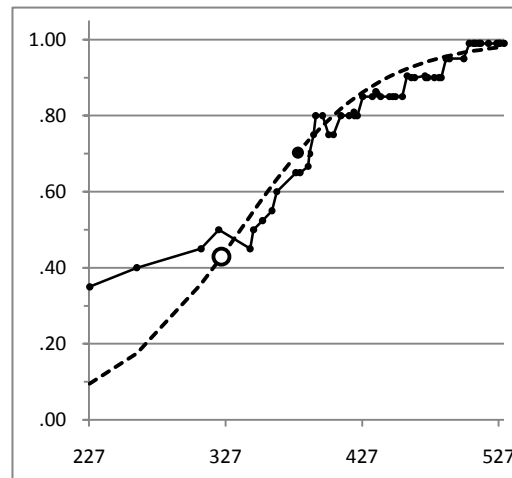
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	374	397	395	448
False Negative	26	65	5	14
True Positive	81	42	43	34
False Positive	27	4	65	12
Sensitivity	.76	.39	.90	.71
specificity	.93	.99	.86	.97
Negative Predictive Power	.94	.86	.99	.97
Positive Predictive Power	.75	.91	.40	.74
Accurate Classification	.90	.86	.86	.95
Kappa	.69	.48	.48	.70

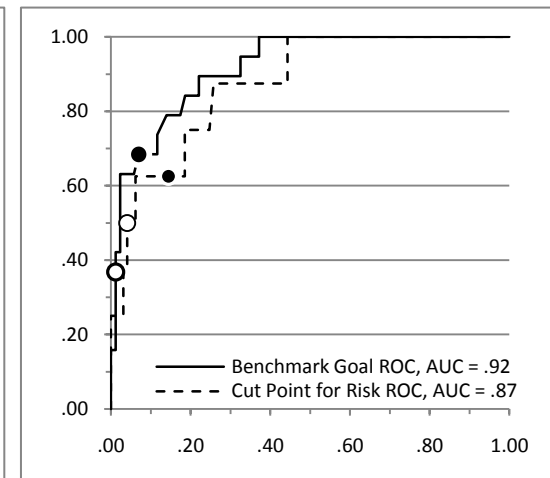
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

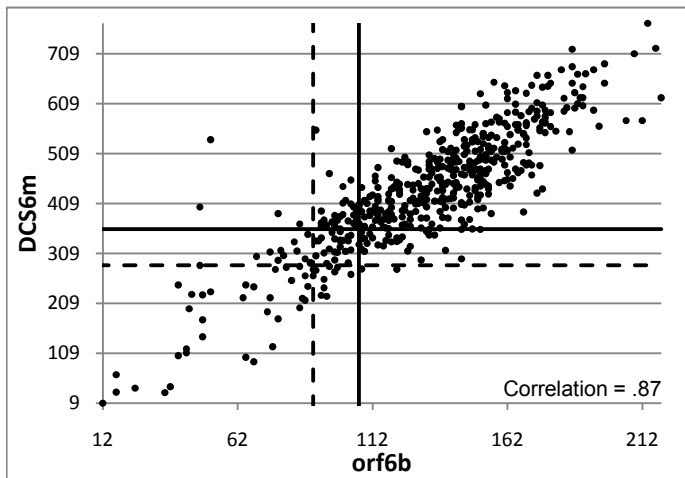
DCS6e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	1	5	80	86	82%
Below Benchmark	3	5	3	11	10%
Well Below Benchmark	4	1	3	8	8%
Marginal Total	8	11	86	105	
Marginal Percent	8%	10%	82%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	13%	45%	93%		

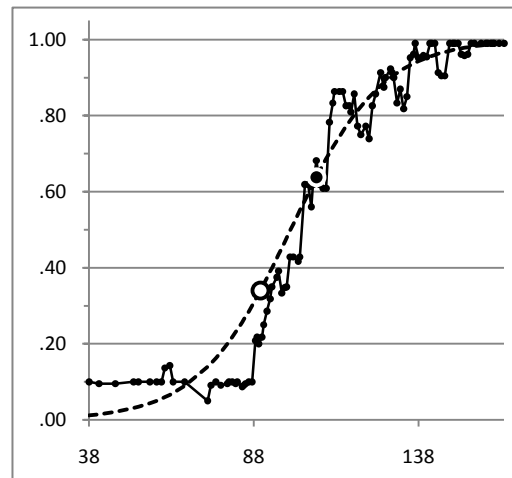
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	80	85	83	93
False Negative	6	12	3	4
True Positive	13	7	5	4
False Positive	6	1	14	4
Sensitivity	.68	.37	.63	.50
specificity	.93	.99	.86	.96
Negative Predictive Power	.93	.88	.97	.96
Positive Predictive Power	.68	.88	.26	.50
Accurate Classification	.89	.88	.84	.92
Kappa	.61	.46	.29	.46

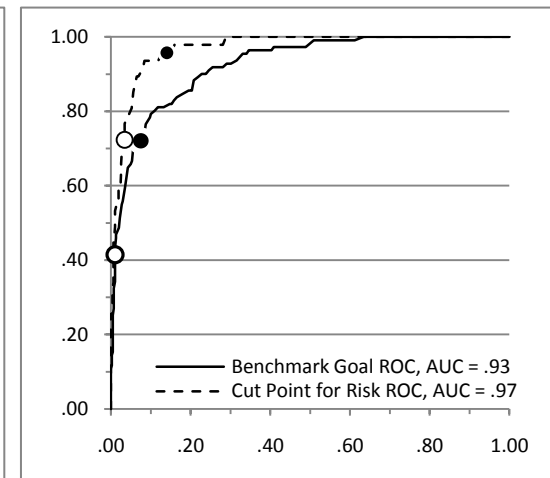
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6b	107	90	DORF Words Correct, Grade 6, Beginning of Year
Outcome	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf6b Screening Decision:

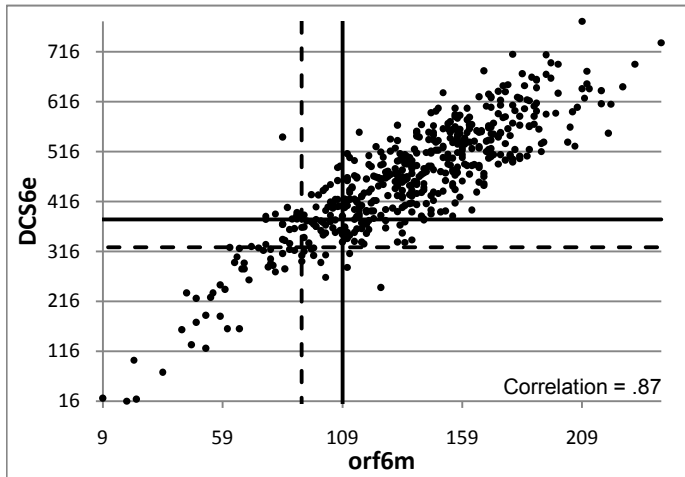
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6m Outcome:					
At or Above Benchmark	4	26	369	399	78%
Below Benchmark	12	23	29	64	13%
Well Below Benchmark	34	11	2	47	9%
Marginal Total	50	60	400	510	
Marginal Percent	10%	12%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	8%	43%	92%		

At or Above Benchmark outcome

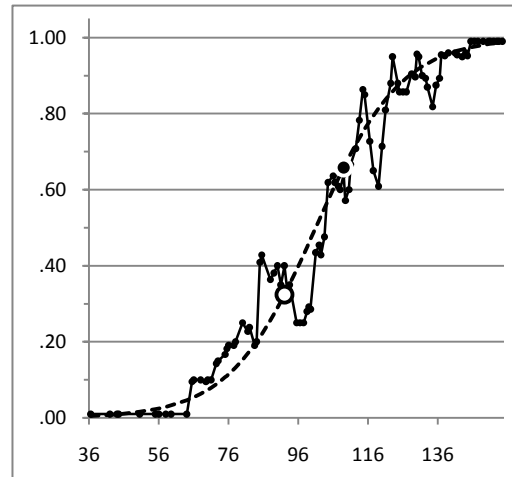
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	369	395	398	447
False Negative	31	65	2	13
True Positive	80	46	45	34
False Positive	30	4	65	16
Sensitivity	.72	.41	.96	.72
specificity	.92	.99	.86	.97
Negative Predictive Power	.92	.86	1.00	.97
Positive Predictive Power	.73	.92	.41	.68
Accurate Classification	.88	.86	.87	.94
Kappa	.65	.50	.51	.67

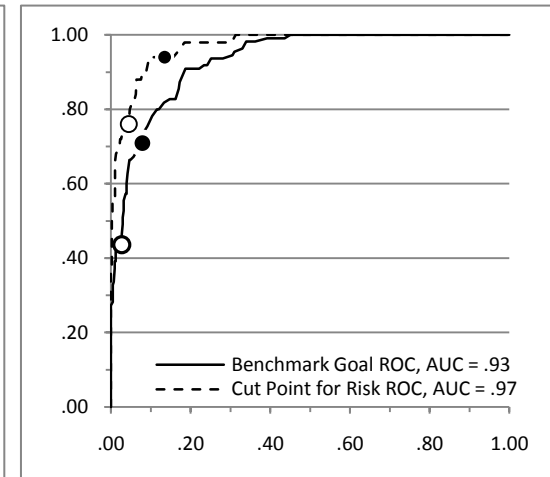
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6m	109	92	DORF Words Correct, Grade 6, Middle of Year
Outcome	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

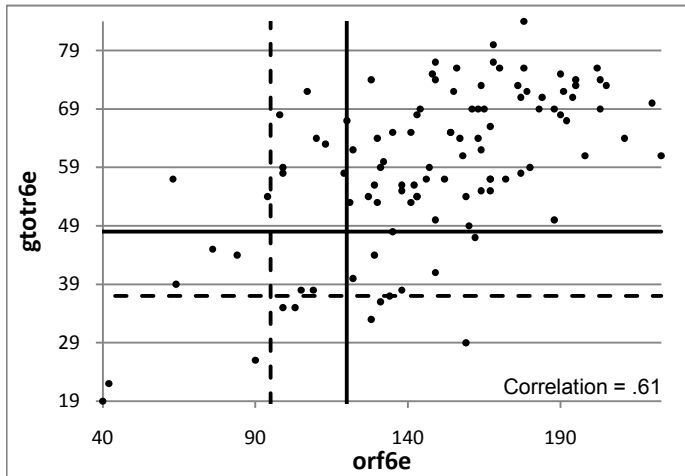
orf6m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6e Outcome:					
At or Above Benchmark	11	21	374	406	79%
Below Benchmark	10	21	29	60	12%
Well Below Benchmark	38	9	3	50	10%
Marginal Total	59	51	406	516	
Marginal Percent	11%	10%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	19%	41%	92%		

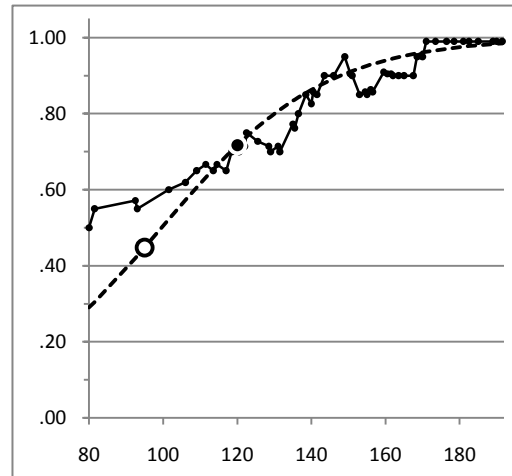
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	374	395	403	445
False Negative	32	62	3	12
True Positive	78	48	47	38
False Positive	32	11	63	21
Sensitivity	.71	.44	.94	.76
specificity	.92	.97	.86	.95
Negative Predictive Power	.92	.86	.99	.97
Positive Predictive Power	.71	.81	.43	.64
Accurate Classification	.88	.86	.87	.94
Kappa	.63	.49	.52	.66

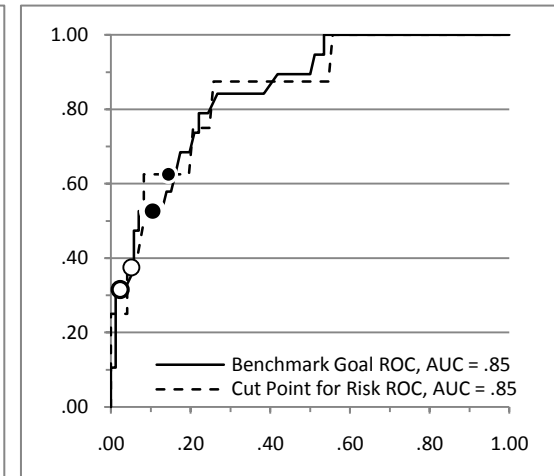
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6e	120	95	DORF Words Correct, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf6e Screening Decision:

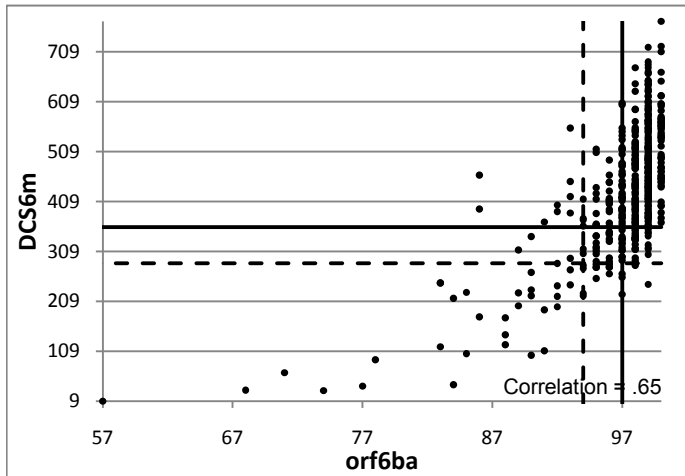
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	2	7	77	86	82%
Below Benchmark	3	2	6	11	10%
Well Below Benchmark	3	2	3	8	8%
Marginal Total	8	11	86	105	
Marginal Percent	8%	10%	82%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	25%	64%	90%		

At or Above Benchmark outcome

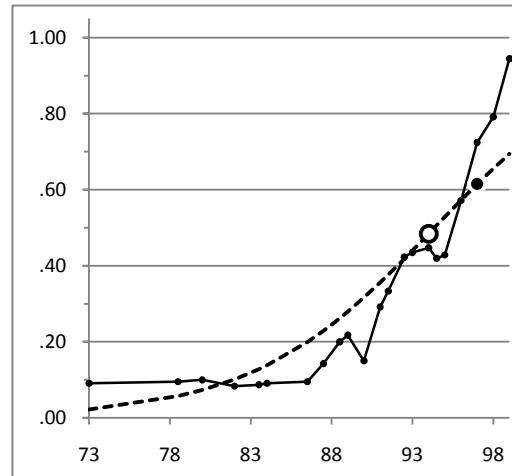
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	77	84	83	92
False Negative	9	13	3	5
True Positive	10	6	5	3
False Positive	9	2	14	5
Sensitivity	.53	.32	.63	.38
specificity	.90	.98	.86	.95
Negative Predictive Power	.90	.87	.97	.95
Positive Predictive Power	.53	.75	.26	.38
Accurate Classification	.83	.86	.84	.90
Kappa	.42	.38	.29	.32

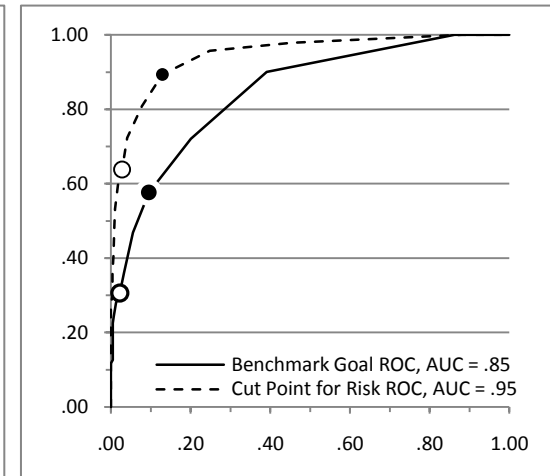
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6ba	97	94	DORF Accuracy, Grade 6, Beginning of Year
Outcome	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

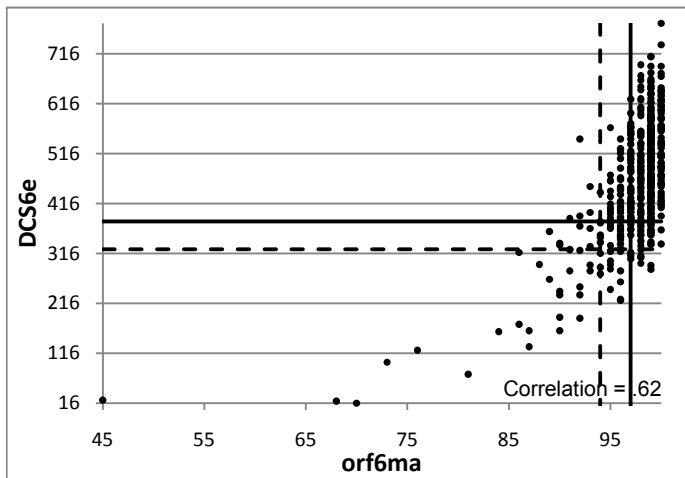
orf6ba Screening Decision:

DCS6m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	9	29	361	399	78%
Below Benchmark	4	18	42	64	13%
Well Below Benchmark	30	12	5	47	9%
Marginal Total	43	59	408	510	
Marginal Percent	8%	12%	80%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	21%	49%	88%		

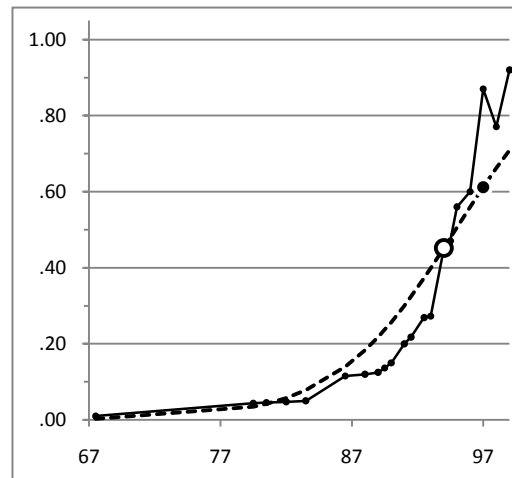
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	361	390	403	450
False Negative	47	77	5	17
True Positive	64	34	42	30
False Positive	38	9	60	13
Sensitivity	.58	.31	.89	.64
specificity	.90	.98	.87	.97
Negative Predictive Power	.88	.84	.99	.96
Positive Predictive Power	.63	.79	.41	.70
Accurate Classification	.83	.83	.87	.94
Kappa	.50	.36	.50	.63

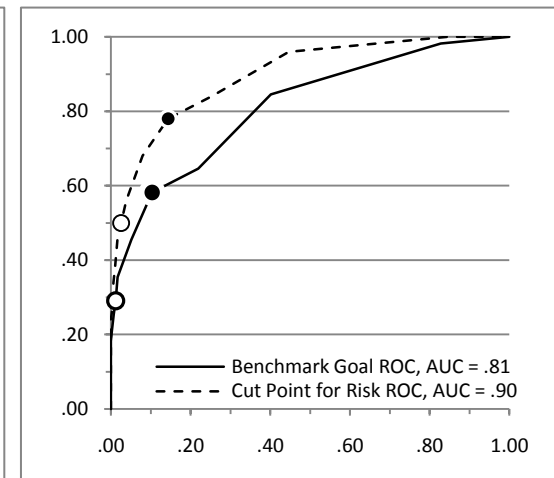
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6ma	97	94	DORF Accuracy, Grade 6, Middle of Year
Outcome	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

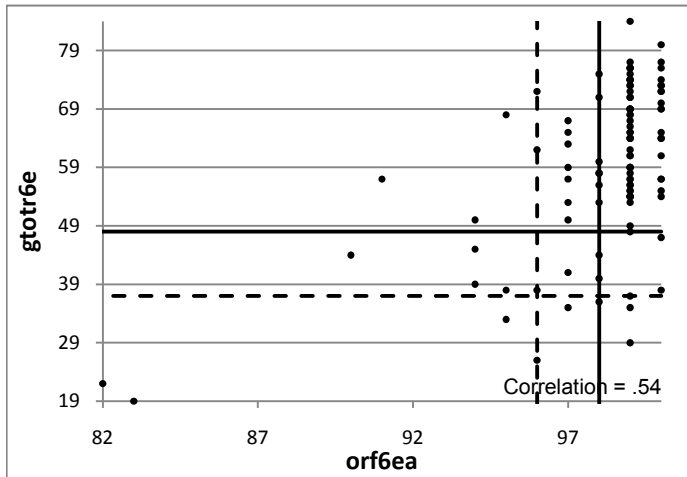
orf6ma Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6e Outcome:					
At or Above Benchmark	5	37	364	406	79%
Below Benchmark	7	18	35	60	12%
Well Below Benchmark	25	14	11	50	10%
Marginal Total	37	69	410	516	
Marginal Percent	7%	13%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	54%	89%		

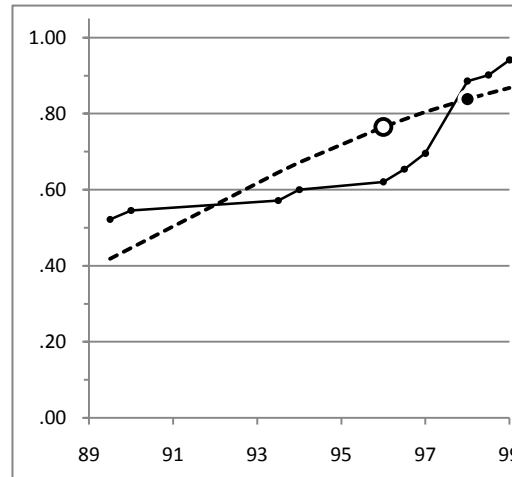
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	364	401	399	454
False Negative	46	78	11	25
True Positive	64	32	39	25
False Positive	42	5	67	12
Sensitivity	.58	.29	.78	.50
specificity	.90	.99	.86	.97
Negative Predictive Power	.89	.84	.97	.95
Positive Predictive Power	.60	.86	.37	.68
Accurate Classification	.83	.84	.85	.93
Kappa	.48	.37	.42	.54

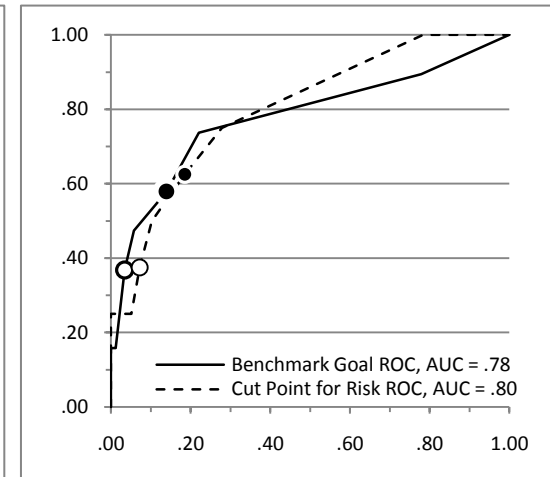
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	orf6ea	98	96	DORF Accuracy, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

orf6ea Screening Decision:

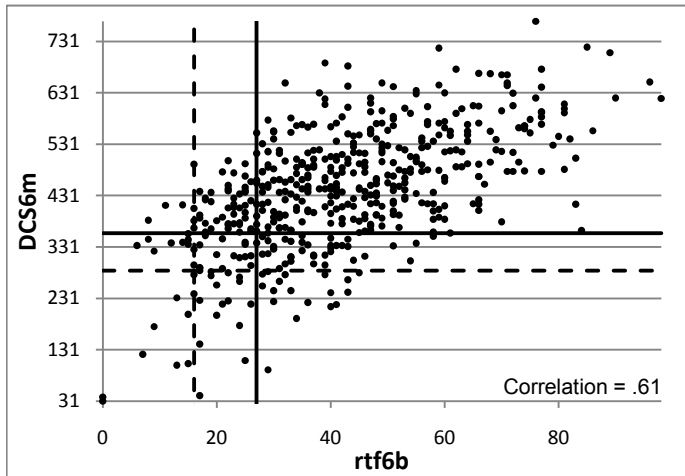
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	3	9	74	86	82%
Below Benchmark	4	2	5	11	10%
Well Below Benchmark	3	2	3	8	8%
Marginal Total	10	13	82	105	
Marginal Percent	10%	12%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	30%	69%	90%		

At or Above Benchmark outcome

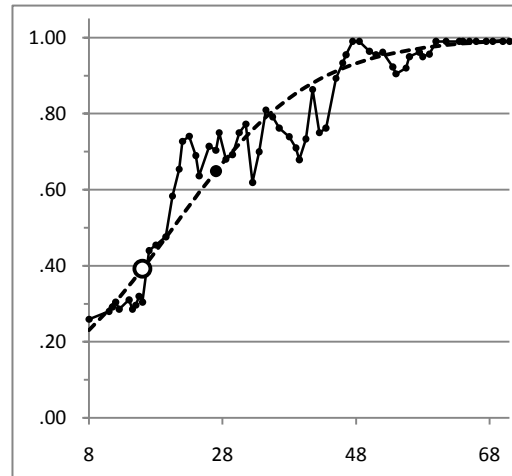
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	74	83	79	90
False Negative	8	12	3	5
True Positive	11	7	5	3
False Positive	12	3	18	7
Sensitivity	.58	.37	.63	.38
specificity	.86	.97	.81	.93
Negative Predictive Power	.90	.87	.96	.95
Positive Predictive Power	.48	.70	.22	.30
Accurate Classification	.81	.86	.80	.89
Kappa	.41	.41	.24	.27

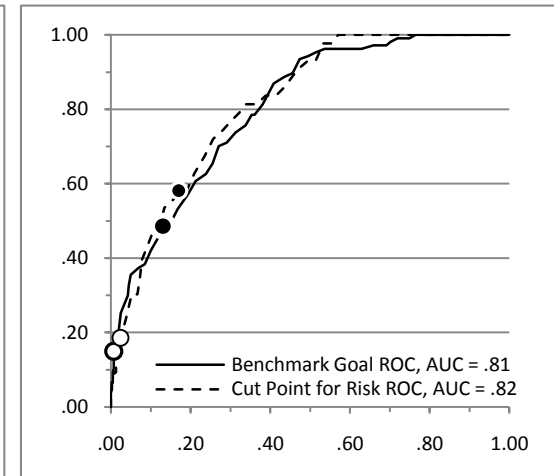
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf6b	27	16	DORF Retell, Grade 6, Beginning of Year
Outcome	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

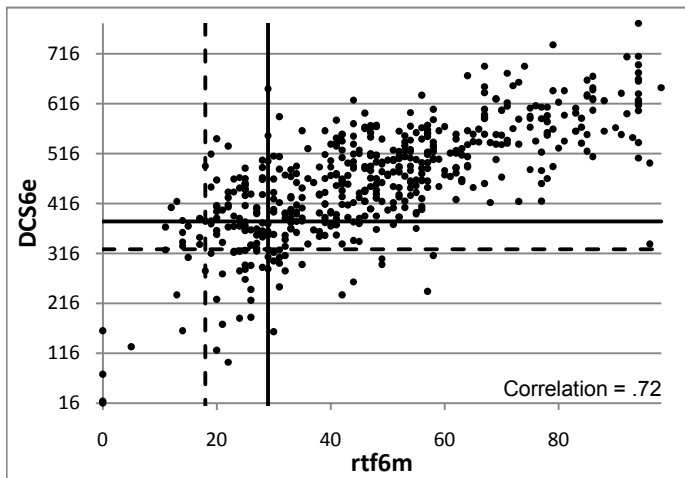
rtf6b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6m Outcome:					
At or Above Benchmark	3	49	347	399	79%
Below Benchmark	8	19	37	64	13%
Well Below Benchmark	8	17	18	43	8%
Marginal Total	19	85	402	506	
Marginal Percent	4%	17%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	16%	58%	86%		

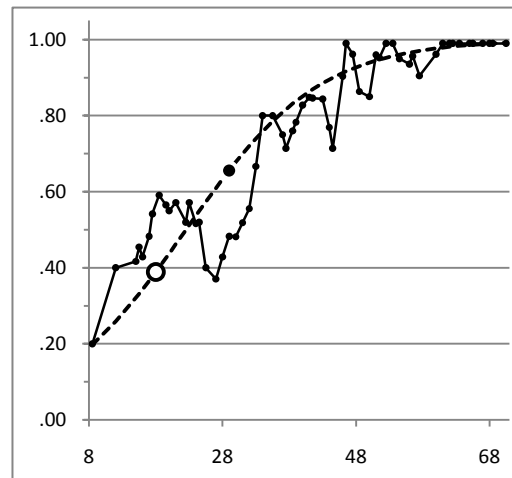
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	347	396	384	452
False Negative	55	91	18	35
True Positive	52	16	25	8
False Positive	52	3	79	11
Sensitivity	.49	.15	.58	.19
specificity	.87	.99	.83	.98
Negative Predictive Power	.86	.81	.96	.93
Positive Predictive Power	.50	.84	.24	.42
Accurate Classification	.79	.81	.81	.91
Kappa	.36	.20	.25	.22

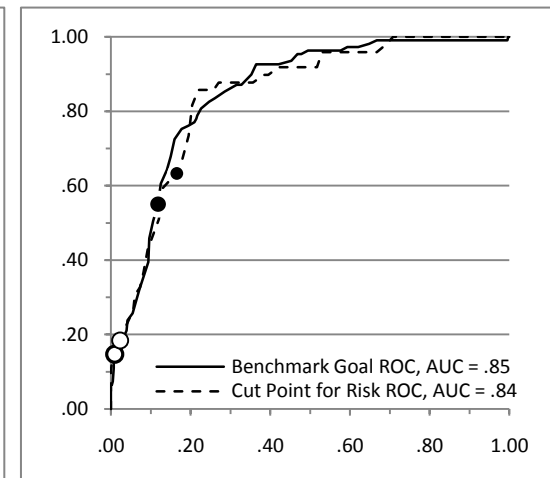
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf6m	29	18	DORF Retell, Grade 6, Middle of Year
Outcome	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rtf6m Screening Decision:

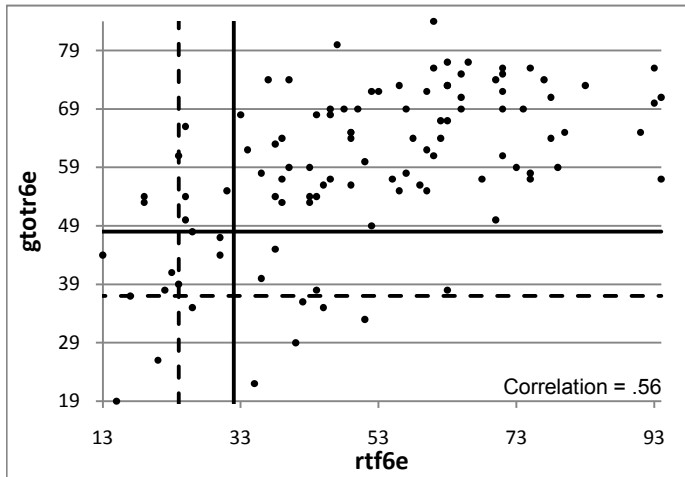
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6e Outcome:					
At or Above Benchmark	4	44	358	406	79%
Below Benchmark	7	22	31	60	12%
Well Below Benchmark	9	22	18	49	10%
Marginal Total	20	88	407	515	
Marginal Percent	4%	17%	79%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	50%	88%		

At or Above Benchmark outcome

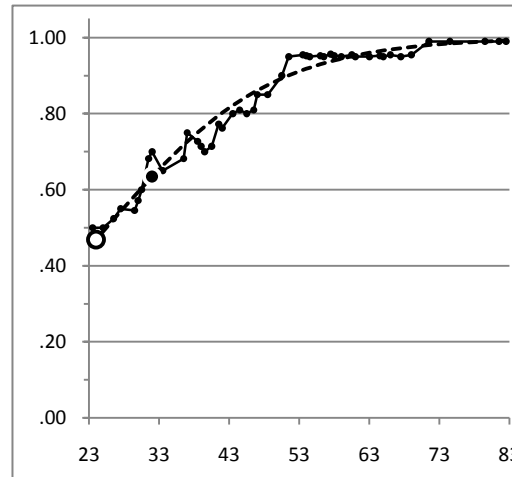
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	358	402	389	455
False Negative	49	93	18	40
True Positive	60	16	31	9
False Positive	48	4	77	11
Sensitivity	.55	.15	.63	.18
specificity	.88	.99	.83	.98
Negative Predictive Power	.88	.81	.96	.92
Positive Predictive Power	.56	.80	.29	.45
Accurate Classification	.81	.81	.82	.90
Kappa	.43	.20	.30	.22

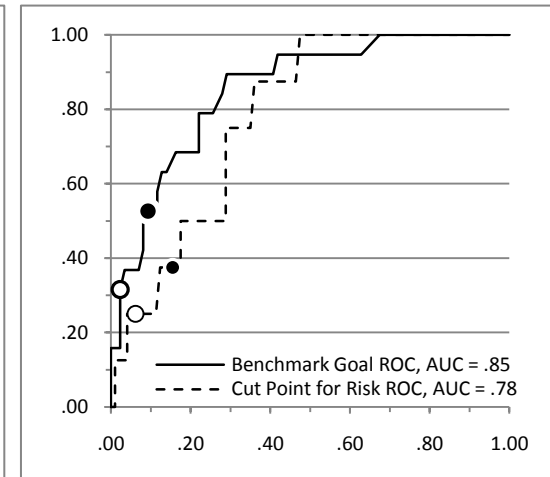
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rtf6e	32	24	DORF Retell, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rtf6e Screening Decision:

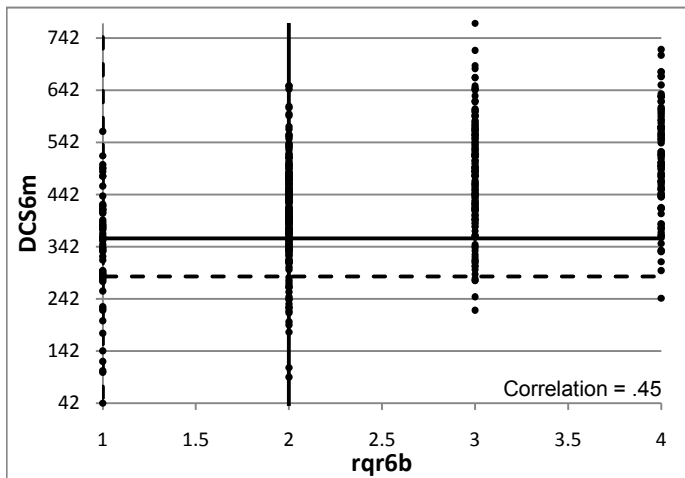
	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	2	6	78	86	82%
Below Benchmark	4	3	4	11	10%
Well Below Benchmark	2	1	5	8	8%
Marginal Total	8	10	87	105	
Marginal Percent	8%	10%	83%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	25%	60%	90%		

At or Above Benchmark outcome

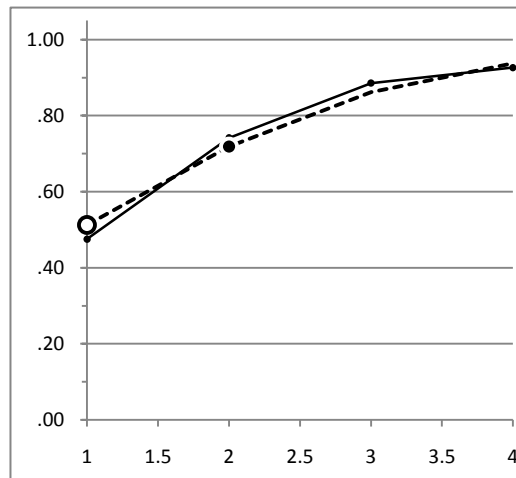
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	78	84	82	91
False Negative	9	13	5	6
True Positive	10	6	3	2
False Positive	8	2	15	6
Sensitivity	.53	.32	.38	.25
specificity	.91	.98	.85	.94
Negative Predictive Power	.90	.87	.94	.94
Positive Predictive Power	.56	.75	.17	.25
Accurate Classification	.84	.86	.81	.89
Kappa	.44	.38	.14	.19

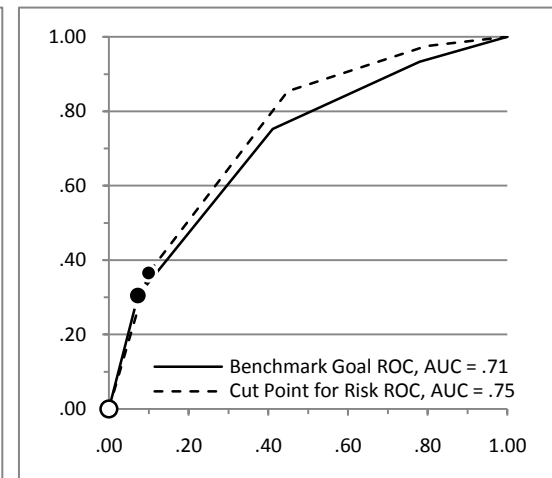
Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	rqr6b	2	1	DORF Retell Quality of Response, Grade 6, Beginning of Year
	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

rqr6b Screening Decision:

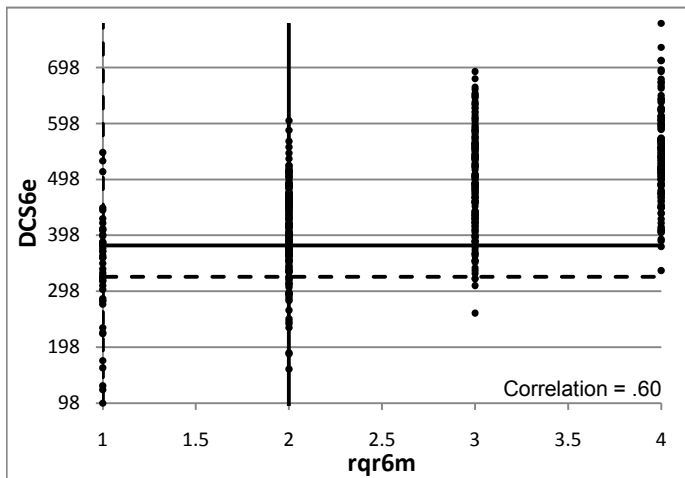
DCS6m Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	0	29	370	399	79%
Below Benchmark	0	17	47	64	13%
Well Below Benchmark	0	15	26	41	8%
Marginal Total	0	61	443	504	
Marginal Percent	0%	12%	88%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	48%	84%		

At or Above Benchmark outcome

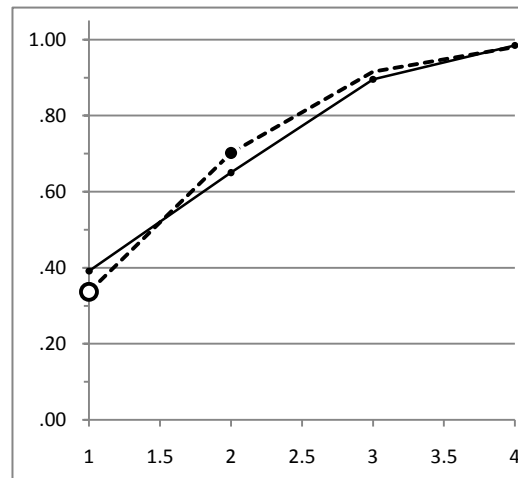
Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	370	399	417	463
False Negative	73	105	26	41
True Positive	32	0	15	0
False Positive	29	0	46	0
Sensitivity	.30	.00	.37	.00
specificity	.93	1.00	.90	1.00
Negative Predictive Power	.84	.79	.94	.92
Positive Predictive Power	.52	--	.25	--
Accurate Classification	.80	.79	.86	.92
Kappa	.27	.00	.22	.00

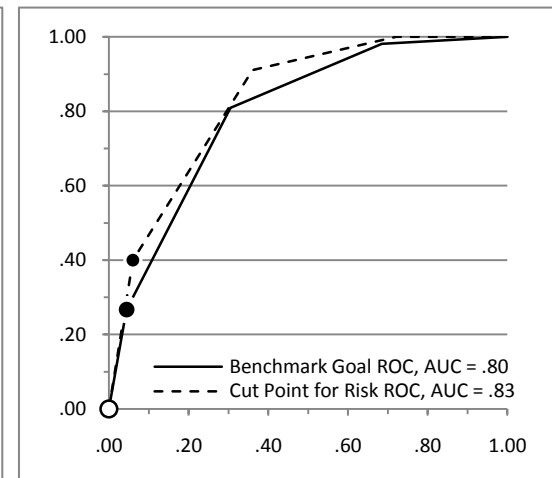
Screening Decision Outcome	Role	Variable	Goal	Cut Point	Description
	Predictor	rqr6m	2	1	DORF Retell Quality of Response, Grade 6, Middle of Year
	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

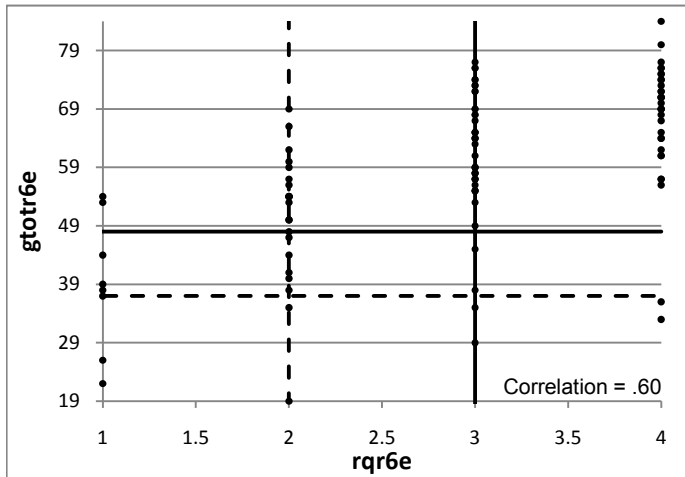
rqr6m Screening Decision:

DCS6e Outcome:	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
At or Above Benchmark	0	18	388	406	79%
Below Benchmark	0	10	50	60	12%
Well Below Benchmark	0	18	27	45	9%
Marginal Total	0	46	465	511	
Marginal Percent	0%	9%	91%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	NA	39%	83%		

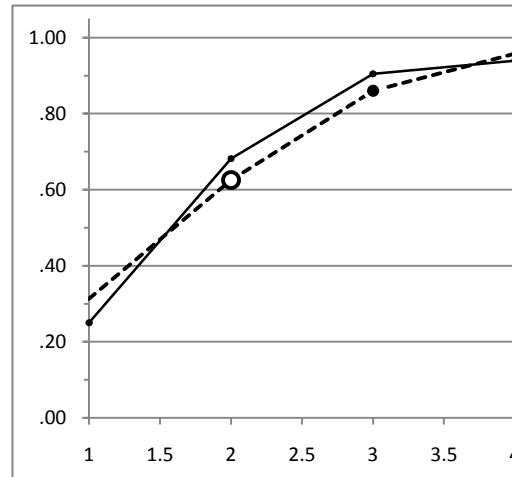
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	388	406	438	466
False Negative	77	105	27	45
True Positive	28	0	18	0
False Positive	18	0	28	0
Sensitivity	.27	.00	.40	.00
specificity	.96	1.00	.94	1.00
Negative Predictive Power	.83	.79	.94	.91
Positive Predictive Power	.61	--	.39	--
Accurate Classification	.81	.79	.89	.91
Kappa	.28	.00	.34	.00

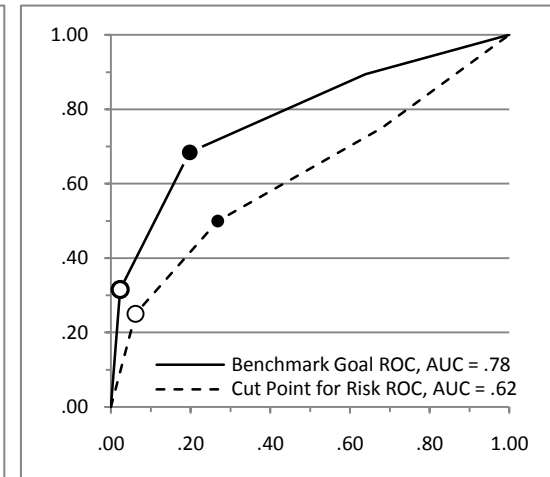
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	rqr6e	3	2	DORF Retell Quality of Response, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

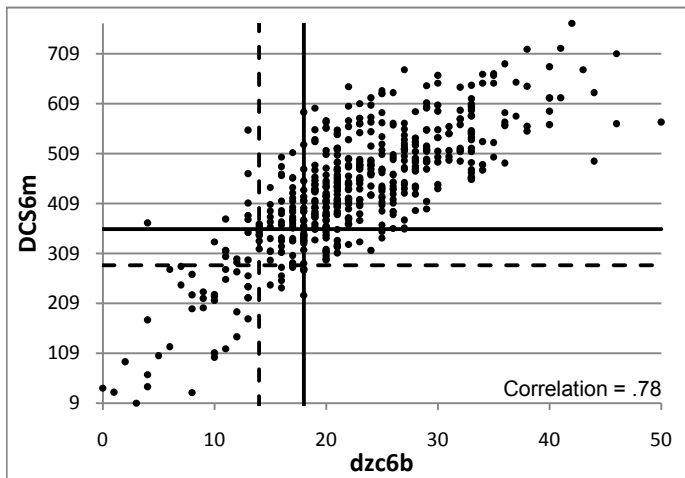
rqr6e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	2	15	69	86	82%
Below Benchmark	4	5	2	11	10%
Well Below Benchmark	2	2	4	8	8%
Marginal Total	8	22	75	105	
Marginal Percent	8%	21%	71%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	25%	68%	92%		

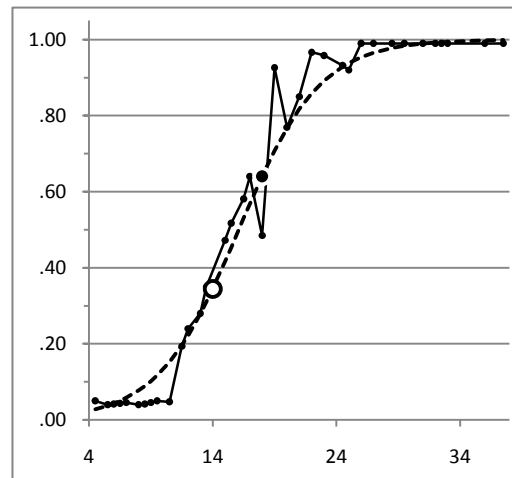
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	69	84	71	91
False Negative	6	13	4	6
True Positive	13	6	4	2
False Positive	17	2	26	6
Sensitivity	.68	.32	.50	.25
specificity	.80	.98	.73	.94
Negative Predictive Power	.92	.87	.95	.94
Positive Predictive Power	.43	.75	.13	.25
Accurate Classification	.78	.86	.71	.89
Kappa	.40	.38	.10	.19

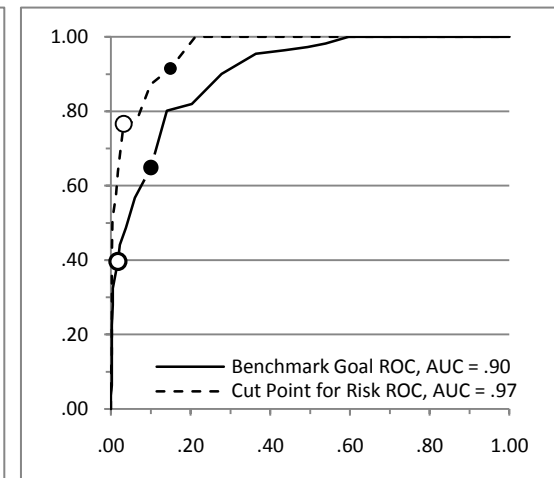
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc6b	18	14	Daze Adjusted Score, Grade 6, Beginning of Year
Outcome	Criterion	DCS6m	358	285	DIBELS Composite Score, Grade 6, Middle of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

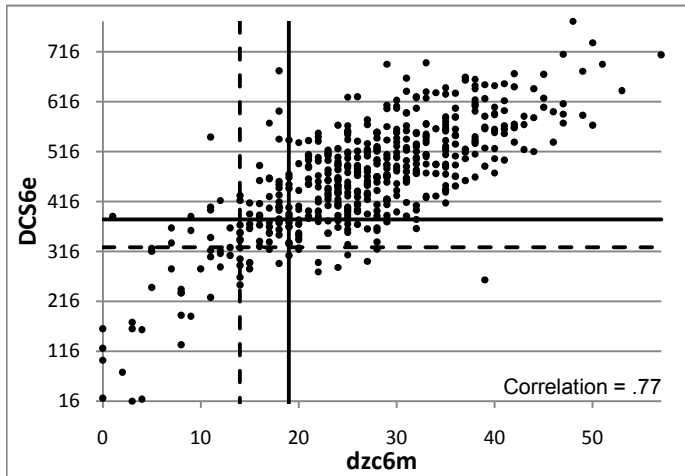
dzc6b Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6m Outcome:					
At or Above Benchmark	7	33	359	399	78%
Below Benchmark	8	21	35	64	13%
Well Below Benchmark	36	7	4	47	9%
Marginal Total	51	61	398	510	
Marginal Percent	10%	12%	78%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	14%	54%	90%		

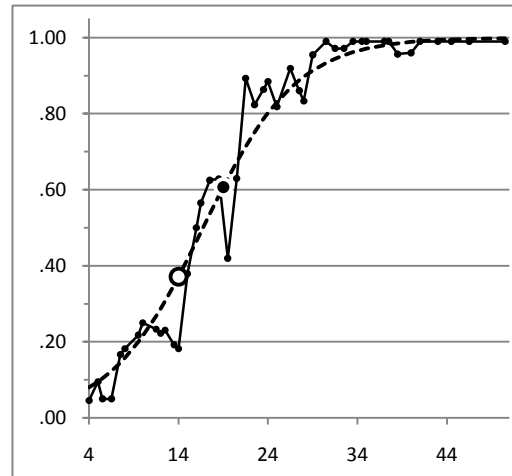
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	359	392	394	448
False Negative	39	67	4	11
True Positive	72	44	43	36
False Positive	40	7	69	15
Sensitivity	.65	.40	.91	.77
specificity	.90	.98	.85	.97
Negative Predictive Power	.90	.85	.99	.98
Positive Predictive Power	.64	.86	.38	.71
Accurate Classification	.85	.85	.86	.95
Kappa	.55	.47	.47	.71

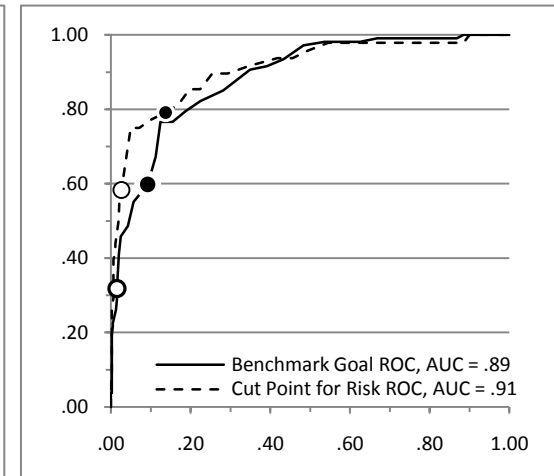
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc6m	19	14	Daze Adjusted Score, Grade 6, Middle of Year
Outcome	Criterion	DCS6e	380	324	DIBELS Composite Score, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

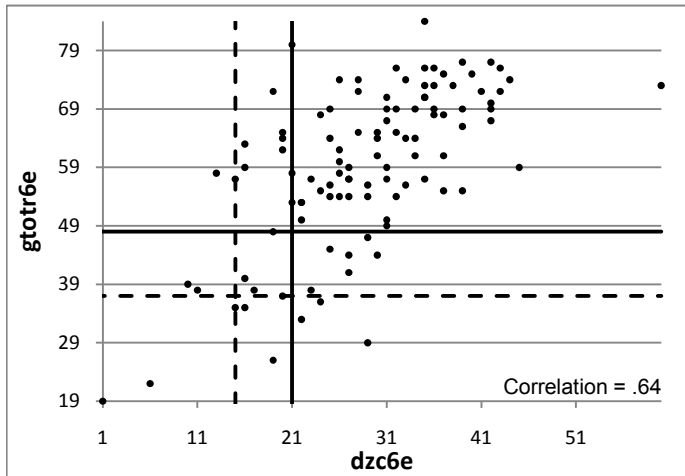
dzc6m Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
DCS6e Outcome:					
At or Above Benchmark	6	31	364	401	79%
Below Benchmark	6	20	33	59	12%
Well Below Benchmark	28	10	10	48	9%
Marginal Total	40	61	407	508	
Marginal Percent	8%	12%	80%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	15%	51%	89%		

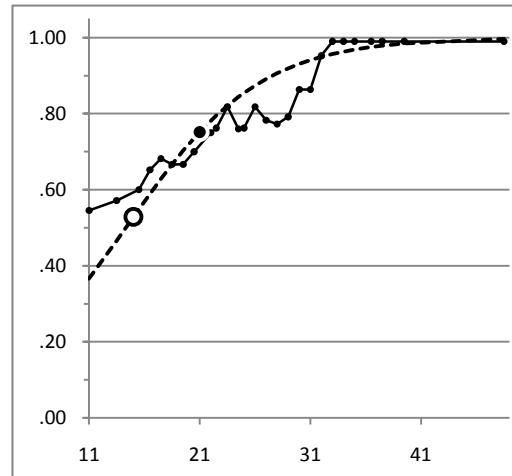
At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	364	395	397	448
False Negative	43	73	10	20
True Positive	64	34	38	28
False Positive	37	6	63	12
Sensitivity	.60	.32	.79	.58
specificity	.91	.99	.86	.97
Negative Predictive Power	.89	.84	.98	.96
Positive Predictive Power	.63	.85	.38	.70
Accurate Classification	.84	.84	.86	.94
Kappa	.52	.39	.44	.60

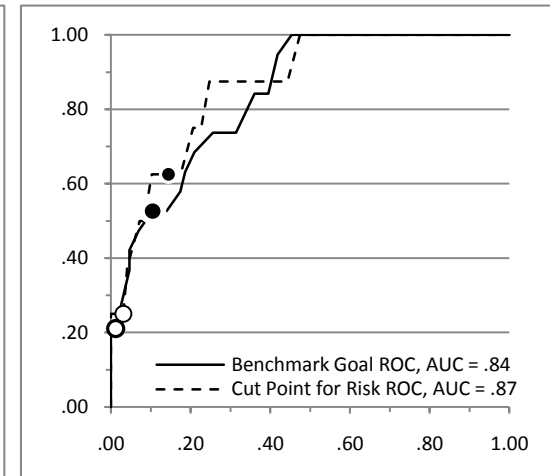
	Role	Variable	Goal	Cut Point	Description
Screening Decision	Predictor	dzc6e	21	15	Daze Adjusted Score, Grade 6, End of Year
Outcome	Criterion	gtotr6e	48	37	GRADE Total Test, Grade 6, End of Year



Scatterplot with benchmark goals (solid lines) and cut points for risk (dashed lines).



Logistic regression with goal (solid dot) and cut point (open dot).



Receiver Operator Characteristic (ROC) curves.

dzc6e Screening Decision:

	Likely to need intensive support	Likely to need strategic support	Likely to need core support	Marginal total	Marginal percent
gtotr6e Outcome:					
At or Above Benchmark	1	8	77	86	82%
Below Benchmark	2	3	6	11	10%
Well Below Benchmark	2	3	3	8	8%
Marginal Total	5	14	86	105	
Marginal Percent	5%	13%	82%		
Odds (conditional percent) of students with screening decision achieving goal (At or Above Benchmark)	20%	57%	90%		

At or Above Benchmark outcome Well Below Benchmark outcome

	Core support decision	Intensive support decision	Core support decision	Intensive support decision
True Negative	77	85	83	94
False Negative	9	15	3	6
True Positive	10	4	5	2
False Positive	9	1	14	3
Sensitivity	.53	.21	.63	.25
specificity	.90	.99	.86	.97
Negative Predictive Power	.90	.85	.97	.94
Positive Predictive Power	.53	.80	.26	.40
Accurate Classification	.83	.85	.84	.91
Kappa	.42	.28	.29	.26

Appendix B: DIBELS Next Composite Score Worksheets.

This appendix includes the worksheets for calculating the DIBELS Next Composite Score from the individual DIBELS Next measures by grade level for each time of year.

K Kindergarten DIBELS® Next Composite Score Worksheet

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____ Class: _____

Beginning of Year Benchmark

FSF Score = _____ [1]

LNF Score = _____ [2]

DIBELS Composite Score (add values 1–2) =

Do not calculate the composite score if any of the values that go into it are missing.

Middle of Year Benchmark

FSF Score = _____ [1]

LNF Score = _____ [2]

PSF Score = _____ [3]

NWF CLS Score = _____ [4]

DIBELS Composite Score (add values 1–4) =

Do not calculate the composite score if any of the values that go into it are missing.

End of Year Benchmark

LNF Score = _____ [1]

PSF Score = _____ [2]

NWF CLS Score = _____ [3]

DIBELS Composite Score (add values 1–3) =

Do not calculate the composite score if any of the values that go into it are missing.

1 First Grade DIBELS® Next Composite Score Worksheet

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning of Year Benchmark

LNF Score = _____ [1]

PSF Score = _____ [2]

NWF CLS Score = _____ [3]

DIBELS Composite Score (add values 1–3) =

Do not calculate the composite score if any of the values that go into it are missing.

Middle of Year	
DORF Accuracy Percent	Accuracy Value
0% – 49%	0
50% – 52%	2
53% – 55%	8
56% – 58%	14
59% – 61%	20
62% – 64%	26
65% – 67%	32
68% – 70%	38
71% – 73%	44
74% – 76%	50
77% – 79%	56
80% – 82%	62
83% – 85%	68
86% – 88%	74
89% – 91%	80
92% – 94%	86
95% – 97%	92
98% – 100%	98

Middle of Year Benchmark

NWF CLS Score = _____ [1]

NWF WWR Score = _____ [2]

DORF Score = _____ [3]

DORF Accuracy Percent: _____ %

100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

Do not calculate the composite score if any of the values that go into it are missing.

End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 64%	0
65% – 66%	3
67% – 68%	9
69% – 70%	15
71% – 72%	21
73% – 74%	27
75% – 76%	33
77% – 78%	39
79% – 80%	45
81% – 82%	51
83% – 84%	57
85% – 86%	63
87% – 88%	69
89% – 90%	75
91% – 92%	81
93% – 94%	87
95% – 96%	93
97% – 98%	99
99% – 100%	105

End of Year Benchmark

NWF WWR Score _____ x 2 = _____ [1]

DORF Score = _____ [2]

DORF Accuracy Percent: _____ %

100 x (Words Correct / (Words Correct + Errors))

Accuracy Value from Table = _____ [3]

DIBELS Composite Score (add values 1–3) =

Do not calculate the composite score if any of the values that go into it are missing.

2 Second Grade DIBELS® Next Composite Score Worksheet

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning of Year	
DORF Accuracy Percent	Accuracy Value
0% – 64%	0
65% – 66%	3
67% – 68%	9
69% – 70%	15
71% – 72%	21
73% – 74%	27
75% – 76%	33
77% – 78%	39
79% – 80%	45
81% – 82%	51
83% – 84%	57
85% – 86%	63
87% – 88%	69
89% – 90%	75
91% – 92%	81
93% – 94%	87
95% – 96%	93
97% – 98%	99
99% – 100%	105

Beginning of Year Benchmark

NWF WWR Score _____ $\times 2$ = _____ [1]

DORF Score = _____ [2]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [3]

DIBELS Composite Score (add values 1–3) =

Do not calculate the composite score if any of the values that go into it are missing.

Middle and End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Middle of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ $\times 2$ = _____ [2]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [3]

DIBELS Composite Score (add values 1–3) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

End of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ $\times 2$ = _____ [2]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [3]

DIBELS Composite Score (add values 1–3) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

3 Third Grade DIBELS® Next Composite Score Worksheet

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning, Middle and End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

Middle of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

End of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

4 Fourth Grade DIBELS® Next Composite Score Worksheet

325

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning, Middle and End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

Middle of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

End of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

5 Fifth Grade DIBELS® Next Composite Score Worksheet

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning, Middle and End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

Middle of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

End of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

6 Sixth Grade DIBELS® Next Composite Score Worksheet

327

© Dynamic Measurement Group, Inc. / August 31, 2010

The DIBELS Composite Score is used to interpret student results for DIBELS Next. Most data management services will calculate the composite score for you. If you do not use a data management service or if your data management service does not calculate it, you can use this worksheet to calculate the composite score.

Name: _____

Class: _____

Beginning, Middle and End of Year	
DORF Accuracy Percent	Accuracy Value
0% – 85%	0
86%	8
87%	16
88%	24
89%	32
90%	40
91%	48
92%	56
93%	64
94%	72
95%	80
96%	88
97%	96
98%	104
99%	112
100%	120

Beginning of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

Middle of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

End of Year Benchmark

DORF Score = _____ [1]

Retell Score _____ x 2 = _____ [2]

Daze Adjusted Score _____ x 4 = _____ [3]

DORF Accuracy Percent: _____ %
 $100 \times (\text{Words Correct} / (\text{Words Correct} + \text{Errors}))$

Accuracy Value from Table = _____ [4]

DIBELS Composite Score (add values 1–4) =

*Do not calculate the composite score if any of the values are missing
 (if DORF is below 40 and Retell is not administered, use 0 for Retell only on these worksheets).*

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Badian, N.A. (1995). Predicting reading ability over the long term: The changing role of letter naming, phonological awareness and orthographic processing. *Annals of Dyslexia*, 45, 79–96.
- Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross-linguistic study of early literacy acquisition. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 145–162). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cummings, K. D., Kaminski, R. A., Good, R. H., & O’Neil, M. E. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of First Sound Fluency. *Assessment for Effective Intervention*, 36(2), 94–106.
doi:10.1177/1534508410392209
- Daane, M.C., Campbell, J.R., Grigg, W.S., Goodman, M.J., & Oranje, A. (2005). *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* (NCES 2006–469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office. Available
<http://nces.ed.gov/nationsreportcard/pdf/studies/2006469.pdf>. Accessed 6/22/2010.
- Dana, J. (2010). What makes improper linear models tick? In J. I. Krueger (ed.) *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 71-90). New York: Taylor & Francis.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.

- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1–17). New York: Guilford Press.
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.
- Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children*, 19(8), 1–15.
- Dewey, E. N., Latimer, R. J., Kaminski, R. A., & Good, R. H. (2011). *DIBELS Next Development: Findings from Beta 2 Validation Study* (Tech. Report No. 10). Eugene, OR: Dynamic Measurement Group. Available: <http://dibels.org/>.
- Dewitz, P., & Dewitz, P. K. (2003). They can read the words, but they can't understand: Refining comprehension assessment. *Reading Teacher*, 56, 422–435.
- Ehri, L.C. (1983). A critique of five studies related to letter-name knowledge and learning to read. In L. Gentile, M. Kamil, & J. Blanchard (Eds.), *Reading research revisited* (pp. 143–153). Columbus, OH: C.E. Merrill.
- Fuchs, D., & Fuchs, L.S. (2006). Introduction to responsiveness-to-intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41, 92–99. doi:10.1598/RRQ.41.1.4
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57(6), 488–500.
- Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/ prevention model of decision making for early literacy skills. *School Psychology*

Quarterly, 11, 326– 336. doi:10.1037/h0088938

- Good, R. H., Simmons, D. C., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 27, 45–56.
- Haynes, S. N., Richard, D. R., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238–247.
- 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>
- Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, 58(8), 702–714.
- Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub.L.No.108–446, 118 Stat. 2647 (2004), [Amending 20 U.S.C. § § 1400 et seq.].
- Kame'enui, E. J., Carnine, D. W., Dixon, R. C., Simmons, D. C., & Coyne, M. D. (2002). *Effective teaching strategies that accommodate diverse learners* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- Kaminski, R. A. (1992). *Assessment for the primary prevention of early academic problems: Utility of curriculum-based measurement prereading tasks*. Unpublished doctoral dissertation. University of Oregon, Eugene.
- Kaminski, R.A., Baker, S.K., Chard, D., Clarke, B., Smith, S. (2006). *Final report: Reliability, Validity, and Sensitivity of Houghton Mifflin Early Growth Indicators* (Technical Report). Eugene, OR: Dynamic Measurement Group and Pacific Institutes for Research.
- Kaminski, R.A. & Cummings, K. D. (2007). *Assessment for learning: Using general outcomes*

- measures. *Threshold*, Winter, 2007, 26–28. Available <http://ciconline.org/threshold>.
- Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for formative assessment and evaluation. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-V*, (pp. 1181–1204). Bethesda, MD: National Association of School Psychologists.
- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215–227.
- Kaminski, R. A., & Good, R. H., III. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 113–142). New York: Guilford Press.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychology Review*, 95, 492–527. doi:10.1037/0033-295X.95.4.492
- Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13, 123–46.
- Mann, V.A., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing: An Interdisciplinary Journal*, 15, 653–682.
- Messick, S. (1993). Validity. In R. L. Linn (Ed.). *Educational measurement* (pp. 105–146). Phoenix, AZ: Oryx Press.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Reports of the subgroups. Washington, DC: National Institute of Child Health and Human

Development.

National Research Council. (1998). *Preventing reading difficulties in young children*.

Washington, DC: National Academy Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS Next Oral Reading Fluency Readability Study* (Tech. Report No. 7). Eugene, OR: Dynamic Measurement Group.
Available: <http://dibels.org/>.

Powell-Smith, K. A., Good, R. H., Latimer, R. J., Dewey, E. N., Kaminski, R. A., & Cummings, K. D. (2011). *DIBELS Next Benchmark Goals Study* (Tech. Report No. 11). Eugene, OR: Dynamic Measurement Group. Available: <http://dibels.org/>.

Rathvon, N., (2004). *Early reading assessment: A practitioner's handbook*. New York: Guilford.
Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education* (Tenth Edition). Boston: Houghton Mifflin Company.

SAS Institute Inc., SAS 9.1.3 Help and Documentation, Cary, NC: SAS Institute Inc., 2002–2004.

SAS Institute Inc., SAS OnlineDoc 9.1.3, Cary, NC: SAS Institute Inc., 2002–2005.

Shanahan, T. (2005). *Review of the DIBELS: Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition*. In R. A. Spies, B. S. Plake, & L. L. Murphy (Eds.), *The sixteenth mental measurements yearbook* (pp. 308–310). Lincoln, NE: The Buros Institute of Mental Measurements.

Shinn, M. R. (1995). Best practices in using curriculum-based measurement in a problem-solving model. In J. G. A. Thomas (Ed.), *Best practices in school psychology III* (Vol. 3, pp. 671–

- 697). Silver Springs, MD: National Association of School Psychologists.
- Shuttleworth, Martyn (2009). Convergent Validity and Discriminant Validity. Retrieved Nov 16, 2010 from Experiment Resources: <http://www.experiment-resources.com/convergent-validity.html>
- Simmons, D. C., & Kame'enui, E. J. (Eds.). (1998). *What reading research tells us about children with diverse learning needs: Bases and basics*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Tilly, W. D., III, (2008). The evolution of school psychology to science-based practice. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 18–32). Washington, DC: National Association of School Psychologists.
- Torgesen, J. K., Wagner, R. K., Rashotte, C.A., Rose, E., Lindamood, P., Conway, T. , & Garvin, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579–593. doi:10.1037/0022- 0663.91.4.579
- U.S. Dept. of Education, National Center for Education Statistics, CCD (2008-09) & PSS (2007–08). Demographic Information for participating school districts. Retrieved February 7, 2011, from <http://nces.ed.gov/>.
- Walsh, D.J., Price, G.G., & Gillingham, M.G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly, 23*, 108–122.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*(2), 159–177. doi:10.2307/747800
- Zeno, S.M., Ivens, S.H., Millard, R.T., Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.