

Considerations in Selecting Benchmark Goals and Cut Points for Risk

Roland H. Good III, Ph.D.
Ruth A. Kaminski, Ph.D.
Dynamic Measurement Group, Inc.

©2012, Dynamic Measurement Group, Inc. 9/4/12
DIBELS®, DIBELS Next®, and DIBELSnet® and are registered trademarks of Dynamic Measurement Group, Inc.

Introduction

- Independent research and opinion on DIBELS Next® is valued – We attend to all research and improve the measures and decision rules when indicated.
- The Center on Teaching and Learning (CTL) Data System and research are completely independent from the authors of DIBELS® and Dynamic Measurement Group.
- **We recommend use of the official DIBELS Next Benchmark Goals.**
- We do not endorse the use of the CTL proposed benchmark goals.

Official DIBELS Next® Benchmark Goals and Cut Points for Risk

- Official DIBELS Next benchmark goals and cut points for risk are available at:
 - <http://dibels.org/papers/DIBELSNextBenchmarkGoals.pdf>
- A detailed description of the design and development of the official DIBELS Next benchmark goals and cut points for risk is provided in the DIBELS Next Technical Manual available on the DIBELS Next download page at:
 - <https://dibels.org/next/index.php>

The screenshot displays the 'UO DIBELS Data System' interface. At the top, there is a navigation bar with 'CTL' and 'UO DIBELS Data System' logos, along with a search bar and user information. Below this is a green navigation menu with links for Home, Administration, Data Entry, Reports, Training, Research, About, and Help. The main content area is titled 'Distribution Report' and contains several form fields: Report (PDF), Scope (School), Year (2011-2012), Grade (Second Grade), District (XXXXXXXXXXXXXXXXXXXX), School (XXXXXXXXXXXXXXXXXXXX), and Assessment (DIBELS Next). Below these fields, there is a section for 'Select assessment period and need for support to continue:'. This section includes 'Assessment Period' (Beginning) and 'Need for Support' (Former Goals). A dropdown menu is open for 'Need for Support', showing three options: 'Former Goals', '-- Select NFS -- Recommended Goals', and 'Former Goals'. Two callout boxes are overlaid on the image: a red box with a red arrow pointing to the first 'Former Goals' option, containing the text 'CTL Proposed Goals - Not Endorsed by the authors of DIBELS'; and a blue box with a blue arrow pointing to the '-- Select NFS -- Recommended Goals' option, containing the text 'Official DIBELS Next Benchmark Goals - Recommended by the authors of DIBELS'. The footer of the page contains 'CTL Links', 'DIBELS Data System Links', and 'Contact Us' information.

Important Considerations

1. **The official DIBELS Next benchmark goals are developed and validated for educational decision making.** CTL proposed goals are based on sensitivity and specificity which are designed for medical screening for existing conditions (e.g., tuberculosis, strep, HIV, cervical cancer).
2. **The official DIBELS Next benchmark goals are designed to work well with different student groups and a broad range of reading outcome measures.** The normative sample used for the CTL proposed goals is not representative of the academic skills of students in the criterion measure norms.
3. **The official DIBELS Next benchmark goals work well in practice for educational decisions.** The CTL proposed goals over-identify students as needing support, and many of those identified achieve subsequent goals.
4. **The DIBELS Next Composite is a valuable predictor and outcome measure.** The CTL proposed goals disregard the importance and value of the composite

1. DIBELS Next® and Educational Decisions

- To evaluate screening tools in education, we recommend using the likelihood or odds of achieving important educational outcomes because:
 - There is intervening instructional support and intervention occurring between the screening assessment and the outcome. *When there is instruction and intervention between the beginning of year screening and the end of year outcome, the labels of “True Positive” and “False Negative” on which sensitivity is based are not meaningful.*
 - Sensitivity indicates the degree to which the assessment captures an existing condition (i.e., a true positive).

6

Official DIBELS Next® Benchmark Goals for Educational Decisions

- Primary design specifications for benchmark goals were based on the odds of achieving subsequent benchmark goals and the student's likely need for support to make adequate progress.
 - ▶ **At or Above Benchmark:** Odds are generally 80% to 90% of achieving subsequent benchmark goals and important reading outcomes. **Students scoring at or above benchmark are likely to make adequate progress with effective core instruction.**
 - ▶ **Below Benchmark:** Odds are generally 40% to 60% of achieving subsequent benchmark goals and important reading outcomes. **Students scoring below benchmark are likely to need strategic support to make adequate progress.**
 - ▶ **Well Below Benchmark:** Odds are generally 10% to 20% of achieving subsequent benchmark goals and important reading outcomes. **Students scoring well below benchmark are likely to need intensive support to make adequate progress.**

7

Sensitivity and Medical Decisions

- The CTL proposed goals cite Pepe (2003), *The statistical evaluation of medical tests for classification and prediction*, in focusing exclusively on sensitivity and specificity.
- They note that “the sensitivity value represents the proportion of “truly” at-risk students who are correctly identified by the screener as being at risk” (p. 5). *However, end of year outcomes in education are not true or existing at the beginning of the year.*
- Sensitivity and specificity are appropriate for medical decisions like tuberculosis screening (for example) where:
 - a) *Screening and diagnosis occur before treatment or intervention take place.* At screening, the condition is existing: The patient truly has TB or truly does not. No treatment or intervention occurs between screening and the determination of the true state that would change the condition.
 - b) A gold standard diagnosis of the true, existing state is generally agreed upon. We are able to determine with reasonable certainty whether the person truly has TB or not.

Sensitivity in Medical Decisions: Screening for *Existing* Tuberculosis

	Screen: positive for tuberculosis	Screen: negative for tuberculosis
True state (existing): Negative for tuberculosis	False Positive	True Negative
True state (existing): Positive for tuberculosis	True Positive	False Negative

- True positives are accurate screening results: The screening indicated the person had tuberculosis and the person *truly* did have tuberculosis at the time they were screened.
- False negatives are screening errors: The screening indicated the person did not have tuberculosis, but *in truth* the person did have tuberculosis at the time they were screened.

DIBELS Next® and Educational Decisions: Ruining the Prediction

	BOY screen: likely to need additional support	BOY screen: likely to make adequate progress with effective core
EOY outcome: at or above benchmark	Successful intervention or support	Adequate progress
EOY outcome: below benchmark or well below benchmark	Intervention or support was not effective in ruining the prediction	Core instruction was not effective in maintaining adequate progress

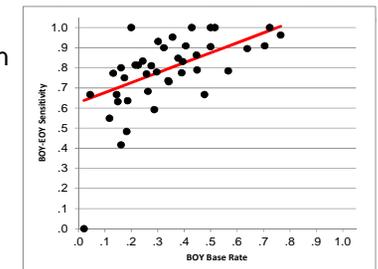
- In education, the EOY outcomes given BOY skills tell us about the effectiveness of our instruction, additional support, or intervention.
- When there is instruction and intervention between the beginning of year screening and the end of year outcome, sensitivity based on “True Positives” and “False Negatives” is not meaningful.

So, Why Does CTL Recommend Sensitivity and Specificity?

- CTL “chose to focus on sensitivity and specificity . . . because they *remain stable indicators regardless of the prevalence of reading difficulties in the population* (Pepe, 2003)” (CTL, 2012, p. 7).
- In fact, measures of sensitivity and specificity for DIBELS Next are not stable.
 - For different districts (populations), indicators of sensitivity range from 0 to 1 – the maximum and minimum possible.
 - Further, the sensitivity indicator appears strongly related to prevalence or base rate for the population.
- We examined the base rate and sensitivity for the DIBELS Next third-grade BOY composite with respect to the third-grade EOY composite for 40 different school districts in the DIBELSnet® data system.
 - Note. Using DIBELS Next Composite Scores and official DIBELS Next Benchmark Goals.

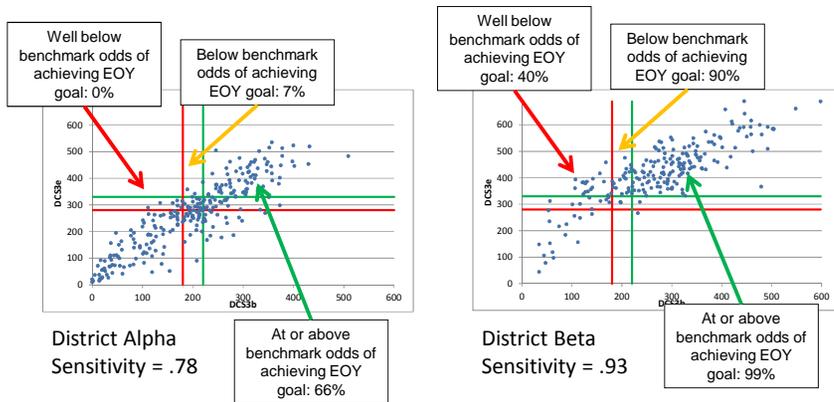
Sensitivity and Specificity of DIBELS Next®

- Sensitivity ranged from 0 to 1.
- BOY base rate and BOY-EOY sensitivity correlated .60, with base rate explaining 37% of the variance in sensitivity, $t(39) = 4.68, p < .0001$.
- Note the district with 0.02 base rate and 0.00 sensitivity (not included in the correlation, the relation is even stronger if it is included).
- **There is no persuasive reason to use sensitivity and specificity in an educational context.**



Note. Using DIBELS Next Composite Scores and official DIBELS Next® Benchmark Goals.

Odds of Achieving Subsequent Goals is Educationally Meaningful



- The differences in odds of achieving the EOY goal are educationally meaningful and interpretable: **The educational support provided by district beta is much more effective in supporting students with the same initial skills to achieve the EOY goal.**
- The difference in sensitivity is not educationally meaningful or interpretable.

Note. Using DIBELS Next® Composite Scores and official DIBELS Next® Benchmark Goals.

No Fate But What We Make

- Districts Alpha and Beta provide a vivid illustration of the fundamental point of this discussion:
 - End of year outcomes are a direct result of what we do between the beginning of year screening and the end of year outcome.
 - End of year outcomes are not pre-existing or true at the time of the screening assessment.
- Odds of achieving outcomes is appropriate for developing and evaluating educational decision models. The odds indicate the amount of support students are likely to need to achieve goals, and they provide a basis for evaluating the support we provide.
- Sensitivity is appropriate for developing and evaluating medical decision models. It indicates the degree to which the assessment identifies a condition that is already present. Sensitivity and specificity do not make sense in educational decision models.

2. Normative Sample Used for the CTL Proposed Goals is Not Representative of the Academic Skills in the Criterion Measure Norms

- CTL identified a representative sample in terms of race, ethnicity, and eligibility for free and reduced price lunch.
- They note, however, that the sample is not representative in terms of academic skills: “across all grades, more than half of our sample is below the 33rd percentile on the SAT10” (p. 10-11).
- When the 50th percentile of the CTL normative sample corresponds to the 33rd percentile of SAT10 normative sample, it indicates that something is amiss. Plausible explanations include:
 - The CTL normative sample is not representative in terms of student achievement.
 - The SAT10 normative sample is not representative in terms of student achievement.
 - Some systematic error in administration and scoring occurred on one or the other measure.

3. Official DIBELS Next® Benchmark Goals are Robust and Valuable in Practice

- We designed the official DIBELS Next benchmark goals and cut points for risk to work with a broad range of students and a wide variety of reading outcome measures.
- The utility of goals can best be examined in a replication with a different sample of students and with a different outcome measure.
- Replication/Comparison School District in California of 29 elementary schools in a large suburban district.
 - 20% of students in the district were ELL students
- Data were available for 22 schools with 72% of elementary students in the district.
 - 46% of students in participating schools reported their race as white
 - 38% identified as Hispanic or Latino students
 - 31% of students qualified for free or reduced lunch

Replication/Comparison Context

- Data to compare DIBELS Next and the California Standards Test were available for
 - 1128 to 1228 students per comparison in second through fifth grades.
 - 553 to 612 students per comparison in sixth grade.
- California Standards Test: The benchmark goal is a standard score of 350 or above, or a performance level of proficient or advanced.
 - “Proficient:** This level represents a solid performance. Students demonstrate a competent and adequate understanding of the knowledge and skills measured by this assessment, at this grade, in this content area.”
 - “Advanced:** This level represents a superior performance. Students demonstrate a comprehensive and complex understanding of the knowledge and skills measured by this assessment, at this grade, in this content area.”

http://star.cde.ca.gov/star2011/help_scoreexplanations.aspx

Are we identifying an appropriate number of students as needing support?

District end of year CST outcomes

- Proficient or advanced: 73%
- Basic or below: 27%

Official DIBELS Next Goals

DIBELS Composite score provides the best predictor of outcomes and best summarizes likely need for support. Across-grade medians at BOY for composite likely need for support:

- core support: 74%
- strategic support: 11%
- intensive support: 12%

CTL Proposed Goals

Focus on individual measures instead of the composite. Need for support “Intensive” if any measure is intensive; “Strategic” if any measure is strategic and no measure is intensive; “Core” if all measures are core. Across-grade medians at BOY:

- core: 24%
- strategic: 18%
- intensive: 58%

Are we identifying an appropriate number of students as needing support?

- Official DIBELS Next goals are identifying an appropriate number of students as likely to need additional support to achieve proficient or above on the California Standards Test.
- CTL proposed goals over identify students as needing strategic support or intensive intervention to achieve the CST goal.
- When too many students are identified for intensive support, our system of intensive support is overwhelmed and we are not able to differentiate instruction at all.**

Goal Utility: Making Educational Decisions

Status	Target odds of achieving outcomes for educational decisions	Likely need for support to achieve outcomes
At or above benchmark	80% to 90%	Core support
Below benchmark	40% to 60%	Strategic support
Well below benchmark	10% to 20%	Intensive support

Official DIBELS Next Goals

Odds of achieving proficient or advanced on the California Standards Test given DIBELS Next likely need for support:

- Core support: 89%
- Strategic support: 36%
- Intensive support: 10%

CTL Proposed Goals

Odds of achieving proficient or advanced on the California Standards Test given CTL proposed need for support:

- Core: 98%
- Strategic: 88%
- Intensive: 59%

Official DIBELS Next® Benchmark Goals are Robust and Valuable in Practice

- **Official DIBELS Next benchmark goals are functioning as designed to inform educational decisions.**
 - Students who are well below benchmark are probably not going to achieve proficient on the CST – unless we provide intensive support and ruin the prediction.
 - For students who are below benchmark, we are not able to make a strong prediction that they will or will not achieve proficient on the CST. They are likely to need strategic support to achieve proficient.
- **CTL proposed goals do not provide helpful guidance in providing additional support.**
 - Students identified as needing strategic support will probably achieve proficient on the CST without additional support.
 - Students identified as needing intensive support will probably achieve proficient on the CST without additional support.

4. The DIBELS Next® Composite is valuable as a predictor and as an outcome measure.

- As a predictor, the DIBELS Next Composite is valuable because students who are at or above benchmark on the DIBELS Composite Score are *reading for meaning* at an *adequate rate* and with a *high degree of accuracy*.
 - Students who are reading as quickly as they can will not score well on the DIBELS Next Composite.
- When multiple measures are administered, it can be confusing to determine an overall level of risk – the DIBELS Next Composite provides an overall indicator of likely need for support.
- These same features make the DIBELS Next Composite valuable as an outcome measure as well.
 - Many group administered reading outcome measures do not assess reading rate or accuracy.
- **Finally, DORF Words Correct alone is a good predictor of outcomes, the DIBELS Composite Score is better.**

Common Core Reading Standards: Foundational Skills (K–5)

Grade 1 to 5 Students

Fluency

4. Read with sufficient accuracy and fluency to support comprehension.
 - a. Read on-level text with purpose and understanding.
 - b. Read on-level prose and poetry orally with accuracy, appropriate rate, and expression on successive readings
 - c. Use context to confirm or self-correct word recognition and understanding, rereading as necessary.

Reading Comprehension Convergence of Information

Beginning of Year Benchmark

DORF Words Correct = _____ [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) =

If DORF is below 40 and Retell is not administered, use 0 for the Retell value only for calculating the DIBELS Composite Score. Do not calculate the composite score if any of the values are missing.

1. Reading at an appropriate rate
2. Reading orally with understanding
3. Reading silently for meaning in context
4. With a high degree of accuracy

Students who are at or above benchmark on the DIBELS® Composite Score are *reading for meaning* at an *adequate rate* and with a *high degree of accuracy*.

Group Reading Assessment and Diagnostic Evaluation

- DIBELS Composite Score explains more variance in reading outcomes than DORF Words Correct alone.
- Median 9% more, range 3% to 17%.
- **DORF Words Correct alone is good, DIBELS Composite Score is better.**

Grade and Time of Year	DORF Words Correct Predicting GRADE	DIBELS Composite Score Predicting GRADE	Additional Variance Explained by DIBELS Composite Score
	Total	Total	Score
Grade 1 Middle of Year	0.64	0.70	8%
Grade 1 End of Year	0.75	0.77	4%
Grade 2 Beginning of Year	0.69	0.75	8%
Grade 2 Middle of Year	0.76	0.80	5%
Grade 2 End of Year	0.73	0.75	3%
Grade 3 Beginning of Year	0.66	0.73	10%
Grade 3 Middle of Year	0.67	0.78	15%
Grade 3 End of Year	0.66	0.75	13%
Grade 4 Beginning of Year	0.76	0.80	5%
Grade 4 Middle of Year	0.76	0.80	6%
Grade 4 End of Year	0.75	0.80	8%
Grade 5 Beginning of Year	0.69	0.76	11%
Grade 5 Middle of Year	0.64	0.76	17%
Grade 5 End of Year	0.66	0.77	17%
Grade 6 Beginning of Year	0.64	0.71	9%
Grade 6 Middle of Year	0.59	0.68	12%
Grade 6 End of Year	0.61	0.73	16%

25

California Standards Test Replication District

- DIBELS Composite Score explains more variance in reading outcomes than DORF Words Correct alone.
- Median 6% more, range 0% to 15%.
- **DORF Words Correct alone is good, DIBELS Composite Score is better.**

Grade and Time of Year	DORF Words Correct Predicting CST	DIBELS Composite Score Predicting CST	Additional Variance Explained by DIBELS Composite Score
	CST	CST	Score
Grade 2 Beginning of Year	.74	.75	1%
Grade 2 Middle of Year	.76	.76	0%
Grade 2 End of Year	.75	.76	2%
Grade 3 Beginning of Year	.68	.71	4%
Grade 3 Middle of Year	.69	.71	3%
Grade 3 End of Year	.69	.73	6%
Grade 4 Beginning of Year	.70	.78	12%
Grade 4 Middle of Year	.72	.77	7%
Grade 4 End of Year	.71	.76	7%
Grade 5 Beginning of Year	.71	.74	4%
Grade 5 Middle of Year	.69	.73	6%
Grade 5 End of Year	.67	.74	10%
Grade 6 Beginning of Year	.67	.74	10%
Grade 6 Middle of Year	.66	.75	13%
Grade 6 End of Year	.63	.74	15%

26

International School in Chile

ITBS Reading Total

- The DIBELS Composite predicts better than DORF Words Correct alone for all comparisons.
- Median 12% more, range 3% to 17%.
- **DORF Words Correct alone is good, DIBELS Composite Score is better.**

Grade	Time of Year	Correlation with ITBS		
		Reading Total		Percent Additional Variance Explained
		DORF Words Correct	DIBELS Composite Score	
Third	BOY	.62	.73	15%
	MOY	.70	.76	9%
	EOY	.71	.79	12%
Fourth	BOY	.67	.72	7%
	MOY	.71	.79	12%
	EOY	.66	.75	13%
Fifth	BOY	.65	.77	17%
	MOY	.72	.77	7%
	EOY	.71	.73	3%
Total	min	.62	.72	3%
	max	.72	.79	17%
	median	.70	.76	12%

27

Summary

- The official DIBELS Next Benchmark Goals provide the best basis for educational decision making.
 1. The official DIBELS Next Benchmark Goals are developed and validated for educational decision making.
 2. The official DIBELS Next Benchmark Goals use procedures that are designed to generalize to different groups of students and many reading outcome measures.
 3. The official DIBELS Next benchmark goals are robust and valuable in practice. They identify an appropriate number of students as likely to need additional support. **When students are identified as likely to need intensive support, the odds are against achieving important goals – unless intensive intervention is provided.**
 4. The DIBELS Next Composite is valuable both as a predictor and as an outcome measure.

See for Yourself

- Schools using DIBELSnet® <https://dibels.net/> can import their own state outcome measure of reading proficiency and examine the utility of the official DIBELS Next benchmark goals for their educational context.
 - Contact us at info@dibels.org if you would like to examine your district and your state outcome measure.
 - DIBELSnet is developed by the authors of DIBELS and provides complete data entry and reporting of DIBELS Next data consistent with the prevention-oriented vision of educational decision making that drove the development of DIBELS.