

# The Robust Beauty of DIBELS® Composite Scores: Homage to Robyn Dawes

Roland H. Good, III, Ruth A. Kaminski, & Kelly A. Powell-Smith, Ph.D. , Dynamic Measurement Group

---

## Introduction and Rationale

Robyn Dawes (1979) describes the robust beauty of improper linear models (ILM) to make important social decisions. With proper linear models (PLM), multiple variables are combined to provide the “best” 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 of 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.

While multiple regression analysis provides the best prediction of the target outcome in terms of most variance explained, “best” could be operationalized in various ways. 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 set of coefficients would best represent the shared variance of multiple variables (factor analysis).

Dawes 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 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 a 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 be a better choice than a PLM. According to Dana (2010) ILM may also be the better option compared to a PLM when predicting beyond a specific sample. Because differences between the sample studied and the population potentially can be a significant source of error, an ILM can provide a better prediction to new and different samples.

In this poster we will demonstrate how a unit-weighted ILM was used as the basis for constructing the *DIBELS* Composite Score in *DIBELS Next*. We illustrate the robust beauty of the *DIBELS* Composite Score for one of the three dimensions identified by Dawes: predicting end-of-year target reading outcomes and predicting end of year generalization reading outcomes.

## Research Question

This study addresses the following primary research question: Does a composite score based upon an improper linear model (ILM) predict a generalization outcome better than the proper linear model (PLM) that best predicts the target outcome?

## Method/Results

Table 1: Correlations of Third Grade, Beginning of Year *DIBELS Next* Measures with End of Year GRADE Total Raw Scores with the Means and Standard Deviations

<i>Third grade, Beginning of year DIBELS Next measure</i>	<i>N with GRADE</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Correlation with Third Grade End of Year GRADE Total Raw Score</i>
DORF Words Correct	184	94.56	36.42	.66
DORF Accuracy	184	94.24	8.17	.68
Accuracy Value (from Table)	184	78.91	35.43	.64
DORF Retell	184	31.42	18.00	.53
Daze	184	11.58	6.32	.67

Table 1 Main Points:

- 1) Beginning of year DORF Words Correct, Accuracy, Retell, and Daze all have strong positive correlations with the target outcome.
- 2) Standard deviations of the measures vary substantially.
- 3) Values from the DORF Accuracy lookup table correlate with the target outcome about the same as DORF Accuracy, and the standard deviation of the accuracy values is approximately equivalent to DORF Words Correct.

Figure 2: Robyn Dawes Unit-Weighted Improper Linear Model (ILM)

<i>Unit-Weighted Improper Linear Model</i>
$ILM = \frac{X_{1i} - \bar{X}_1}{S_1} + \frac{X_{2i} - \bar{X}_2}{S_2} + \frac{X_{3i} - \bar{X}_3}{S_3} + \frac{X_{4i} - \bar{X}_4}{S_4}$
<p>Where: X1 = DORF Words Correct, X2 = DORF Accuracy, X3 = DORF Retell, X4 = Daze</p>
<i>Unit-Weighted Improper Linear Model with Rearranged Terms</i>
$S_1 * ILM = X_{1i} + \frac{S_1}{S_2} X_{2i} + \frac{S_1}{S_3} X_{3i} + \frac{S_1}{S_4} X_{4i} + constant$
<i>For Third Grade, Beginning of Year DIBELS Next</i>
$S_1 * ILM = X_{1i} + 4.46X_{2i} + 2.02X_{3i} + 5.76X_{4i} + constant$
$DCS = X_{1i} + (accuracy\_value) + 2X_{3i} + 4X_{4i}$

Table 2: Comparisons of *DIBELS* Composite Score Predictions with (a) Best Single Measure Predictions and (b) Linear Regression Model Predictions for (a) Target Reading Outcome and (b) Generalization Reading Outcome

	<i>Beginning of year</i>			<i>Middle of year</i>			<i>End of year</i>		
	Best Single Measure	Linear Regression Model	<i>DIBELS</i> Composite Score	Best Single Measure	Linear Regression Model	<i>DIBELS</i> Composite Score	Best Single Measure	Linear Regression Model	<i>DIBELS</i> Composite Score
<b>Kindergarten</b>									
GRADE Total raw score	.52	.53	.50	.47	.51	.48	.40	.41	.37
<i>DIBELS</i> Composite Score	.43	.49	.52	.65	.68	.71			
<b>First Grade</b>									
GRADE Total raw score	.43	.58	.55	.64	.80	.70	.75	.81	.77
<i>DIBELS</i> Composite Score	.71	.73	.73	.83	.84	.89			
<b>Second Grade</b>									
GRADE Total raw score	.69	.79	.75	.76	.83	.80	.73	.78	.75
<i>DIBELS</i> Composite Score	.81	.80	.81	.87	.87	.89			
<b>Third Grade</b>									
GRADE Total raw score	0.66	.78	0.73	0.67	.81	0.78	0.66	.76	0.75
<i>DIBELS</i> Composite Score	0.86	.86	.88	0.86	.84	0.90			
<b>Fourth Grade</b>									
GRADE Total raw score	.76	.79	.80	.76	.79	.80	.75	.80	.80
<i>DIBELS</i> Composite Score	.86	.90	.89	.87	.91	.90			
<b>Fifth Grade</b>									
GRADE Total raw score	.69	.76	.76	.64	.76	.76	.66	.78	.77
<i>DIBELS</i> Composite Score	.84	.85	.86	.86	.87	.90			
<b>Sixth Grade</b>									
GRADE Total raw score	.64	.71	.71	.59	.70	.68	.61	.75	.73
<i>DIBELS</i> Composite Score	.86	.89	.90	.87	.89	.91			

Note. Correlations with GRADE are based on  $n = 103$  to  $219$ . Correlations with end of year *DIBELS* Composite Score based on  $n = 433$  to  $569$ . The end of year GRADE Total Raw Scores are the target reading outcome because each linear regression model was estimated using the GRADE as the dependent variable. The end of year *DIBELS* Composite Score is a generalization outcome because it is a different measure of the same construct.

**Table 2 Main Points:**

- 1) The linear regression model typically provides the best prediction of the target outcome (end of year GRADE Total Raw Score) [illustrated with red circle].
- 2) The *DIBELS* Composite Score typically provides the best prediction of the generalization outcome (end of year *DIBELS* Composite Score) [illustrated with green circle].
- 3) The *DIBELS* Composite Score typically provides the second best prediction of the target outcome, and usually a close second [illustrated with blue circle].

The data for the development of the *DIBELS* Composite Score were from a study designed for the purpose of developing benchmark goals for the *DIBELS Next* assessments.

**Participants**

- Students were recruited from 13 schools in five school districts representing five US regions.
- School districts had a median of 10 years experience using *DIBELS*.
- K–6th grade students participated in *DIBELS Next* assessments ( $n = 3,816$  total; 433 to 569 per grade). The percentage of this sample at benchmark ranged from 65%–79% across grades and times of year.

# 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 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.

## Middle 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.

## End 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.

DIBELS is a registered trademark of Dynamic Measurement Group, Inc. <http://dibels.org/>

- Subsamples of students participated in testing with the Group Reading Assessment and Diagnostic Evaluation (GRADE) (n = 1257 total; 103 to 219 per grade).
- The GRADE sub-sample was 50% female on average across grades.

### Measures

Measures in this study included all DIBELS Next measures and the Group Reading Assessment and Diagnostic Evaluation (GRADE). The GRADE served as our “target” outcome.

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

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

### Group Reading Assessment and Diagnostic Evaluation (GRADE):

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

### Procedures

- All Data were collected during the 2009–2010 school year
- DIBELS Next assessments were administered at regular benchmark intervals.
- GRADE testing was conducted across two to three sessions in the spring. Testing time ranged from 60 to 90 minutes.

- Data for students with missing or duplicate IDs were removed.
- We also removed data for scores that were invalid due to known data collection errors, invalid score ranges, or significant univariate or bivariate outliers.

### Data Analysis

Data analysis proceeded according to the following steps:

- First, we examined correlations between *DIBELS Next* measures and the GRADE Total Test Raw Score and selected the measures with substantial positive correlations.
- Next, we computed the unit-weighting coefficients using the procedures described by Dawes (as illustrated in Figure 2).
- Then, we scaled the resulting composite so that the coefficient for DORF Words Correct would be equal to 1 (as illustrated in Figure 2).
- Next, we specified integer weights for the *DIBELS Composite Score* that would approximate the Dawes improper linear model weights (as illustrated in Figure 2). For DORF Accuracy we used a look-up table to obtain appropriately-scaled values as illustrated in Table 1.
- Finally, we correlated the resulting *DIBELS Composite Scores* with the target outcome variable and with the generalization outcome variable and compared those correlations to the corresponding correlations from the single best *DIBELS* measure in isolation and the correlations from a multiple regression analysis (i.e., proper linear model).

## Conclusions and Discussion

In this study, we formed the *DIBELS Composite* by combining the *DIBELS Next* measures that correlate highly with later outcomes for each grade and time of year and then weighted each measure so **each contributed approximately equally**. The *DIBELS Composite* is highly correlated with target and generalization reading outcomes, and is expected to generalize to a broad range of reading outcomes.

### The Robust Beauty of the *DIBELS Composite Score*

The *DIBELS Composite Score* provides a more complete sample of reading behavior than any **single** measure. For example, beginning of year 3rd grade DORF Words Correct correlates with end of the year GRADE Total Score .66, which is very good. However, beginning of year 3rd grade *DIBELS Composite Score* correlates .73, explaining 10% more variance than DORF alone. The *DIBELS Composite Score* beats the single best *DIBELS Next* measure at almost every grade and time of year.

Importantly, the *DIBELS Composite Score* represents a large, rich, and broad sample of reading behavior. It represents a combining of information from across *DIBELS Next* measures administered at a given time. As such, educators do not need to determine which scores are most important or how to integrate the information. Dawes suggests that “people are bad at integrating information from diverse and incomparable sources” (p. 574). The beauty of the *DIBELS Composite score* is that it allows for easy and meaningful integration of information. The *DIBELS Composite Score* conveys that *all of the aspects of reading proficiency are critical*—a student whose *DIBELS Composite Score* is At or Above Benchmark is reading **accurately**, at an **adequate rate**, and **attending to meaning**.

## References

- 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(7), 571–582.

### Contact information:

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

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

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