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

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

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)

**Unit-Weighted Improper Linear Model**

\[
ILM = \frac{X_{i1} - \bar{X}_1}{s_1} + \frac{X_{i2} - \bar{X}_2}{s_2} + \frac{X_{i3} - \bar{X}_3}{s_3} + \frac{X_{i4} - \bar{X}_4}{s_4}
\]

Where: \(X_1 = \text{DORF Words Correct}, X_2 = \text{DORF Accuracy}, X_3 = \text{DORF Retell}, X_4 = \text{Daze}\)

**Unit-Weighted Improper Linear Model with Rearranged Terms**

\[
s_1*ILM = X_{i1} + \frac{s_1}{s_2}X_{i2} + \frac{s_1}{s_3}X_{i3} + \frac{s_1}{s_4}X_{i4} + \text{constant}
\]

**For Third Grade, Beginning of Year DIBELS Next**

\[
s_1*ILM = X_{i1} + 4.46X_{i2} + 2.02X_{i3} + 5.76X_{i4} + \text{constant}
\]

\[
DCS = X_{i1} + (\text{accuracy\_value}) + 2X_{i3} + 4X_{i4}
\]
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

<table>
<thead>
<tr>
<th></th>
<th>Beginning of year</th>
<th>Middle of year</th>
<th>End of year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Single Measure</td>
<td>Linear Regression Model</td>
<td>DIBELS Composite Score</td>
</tr>
<tr>
<td><strong>Kindergarten</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.52</td>
<td>.53</td>
<td>.50</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.43</td>
<td>.49</td>
<td>.52</td>
</tr>
<tr>
<td><strong>First Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.43</td>
<td>.58</td>
<td>.55</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.71</td>
<td>.73</td>
<td>.73</td>
</tr>
<tr>
<td><strong>Second Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.69</td>
<td>.79</td>
<td>.75</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.81</td>
<td>.80</td>
<td>.81</td>
</tr>
<tr>
<td><strong>Third Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>0.66</td>
<td>.78</td>
<td>.73</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>0.86</td>
<td>.86</td>
<td>.88</td>
</tr>
<tr>
<td><strong>Fourth Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.76</td>
<td>.79</td>
<td>.80</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.86</td>
<td>.90</td>
<td>.89</td>
</tr>
<tr>
<td><strong>Fifth Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.69</td>
<td>.76</td>
<td>.76</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.84</td>
<td>.85</td>
<td>.86</td>
</tr>
<tr>
<td><strong>Sixth Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE Total raw score</td>
<td>.64</td>
<td>.71</td>
<td>.71</td>
</tr>
<tr>
<td>DIBELS Composite Score</td>
<td>.86</td>
<td>.89</td>
<td>.90</td>
</tr>
</tbody>
</table>

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.
- 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

Contact information:
Roland H. Good, III / rhgood@dibels.org
Ruth A. Kaminski / rkamins@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.