

Technical Report #8

Technical Adequacy of Second Grade DIBELS Oral Reading Fluency Passages

Roland Good

Ruth Kaminski

Sylvia Smith

John Bratten

University of Oregon

Good, R. H., Kaminski, R. A., Smith, S. Bratten, J. (in progress). Technical Adequacy of Second Grade DIBELS Oral Reading Fluency Passages. Technical Report. Eugene, OR: University of Oregon.

Grade 2 DORF Passage Analysis

Examination of the passage difficulty levels was done with a sample of 133 children who participated in a 5-week calibration study. Once a week, children were administered 5 DIBELS Oral Reading Fluency (DORF) passages and 1 Test of Reading Fluency (TORF) passage (Children's Educational Services, 1987). Children were administered a total of 25 different DORF passages and 5 TORF passages. The order of administration of the TORF passages was held constant across all children, but the DORF passages were arranged at random into 3 orders of administration.

A first step in the analysis was to compute a growth curve for individual children based on their combined DORF and TORF passages using Hierarchical Linear Modeling (HLM, Bryk, Raudenbush, & Congdon, 1996) procedures. The week in which the passage was administered was the independent variable and the child's number of correct words per minute was the dependent variable for the Level 1, within-child, analysis. For example, for one group of children, "My Dress Up Box" was administered in week 10, for a second group of children it was administered in week 14, and for a third group it was administered in week 5. At Level 2, the between-child analysis, of the HLM, a null model without predictor variables was specified. The intercept of the Level 2 model was specified to be week 0 of the study, the week before the initial assessment. The estimate of the mean intercept was 71.44 and the mean slope of progress was 1.90, and both were significantly different from 0, $t = 22.09$, $p < .01$ and $t = 9.13$, $p < .01$, respectively. The standard deviation of the intercept was 36.91, and the standard deviation of the slope was 1.73. The variability in intercepts and slopes were both significant, $\chi^2(131) = 6287.46$, $p < .01$ and $\chi^2(131) = 274.76$, $p < .01$, respectively. The estimated reliability of the intercept was .98, and the estimated reliability of the slope was .52.

Of primary interest in this analysis was the residual of each passage score from the score predicted by the individual child growth trajectory. To compute a passage residual, the child's predicted score based on the growth trajectory using all of the passages was subtracted from the

actual score for the passage. A positive passage residual was indicative that the child scored better on the passage than would be expected from their overall growth trajectory. The standard deviation of residuals was 12.27.

The mean residual for each passage indicated whether the passage was systematically easier (positive mean residual) or more difficult (negative mean residual) compared to the overall difficulty level of all passages used to estimate the growth curves. The average of the mean residuals of the TORF passages was 9.45, indicating that overall the DORF passages were about 9 words correct per minute more difficult than the TORF passages.

To reference the difficulty level of the DORF passages with respect to the TORF passages, an adjusted mean residual was computed for each passage by subtracting 9.45 from the mean residual of each passage. The average of adjusted mean residuals was then 0 for the TORF passages. The adjusted mean residual for the DORF passages indicates whether the passage was systematically more difficult (negative adjusted mean residual) or easier (positive adjusted mean residual) compared to the average TORF passage. The 9 DORF passages with smallest adjusted mean residuals initially were selected for the DORF Benchmark Assessment Passages. The adjusted mean residuals for the DORF Passages and the TORF passages are reported in Table 1. In this step of the establishment of the DORF Passages, the DORF passages were about 5 words per minute more difficult than the TORF passages.

In the second phase of the analysis, the readability of all passages was estimated using the Micro Power & Light readability software (2000a; 2000b). All readability estimates were computed, including Dale-Chall, Flesch, FOG, Powers*, SMOG, FORCAST, Frye, and Spache. The readabilities of the passages were entered as independent variables in a stepwise regression analysis with the adjusted mean residual as the dependent variable. Passage was the unit of analysis. The Spache readability was entered first in the analysis and accounted for 30% of the variance in adjusted mean residual. No other variables met the criteria for inclusion once the Spache readability had been entered. Based on the regression analysis, the best estimate of the adjusted mean residual from the readability formulas was, Predicted Adjusted Mean Residual =

21.719 - 12.431 * Spache. This prediction is far from exact. It is interesting to note that, using the prediction, the TORF passages were predicted to be somewhat more difficult than they actually were.

Informed by the pattern of adjusted mean residuals, minor revisions were made to the passages to make their difficulty level more consistent with the TORF passages. Difficult words were replaced with more common words, and complex sentence structures were revised. The readability estimates and predicted adjusted mean residuals for the revised passages are reported in Table 1. Based on the prediction, the revised DORF Assessment passages are predicted to be about 1 correct word per minute more difficult than the TORF. In addition, the range in readability estimates for the revised DORF passages compares favorably with the TORF passages. Of course, the estimated readability of the revised DORF passages must be examined empirically before firm conclusions can be drawn.

Table 1

Comparison of DORF and TORF Passage Difficulty Levels for Second Grade Students

Passage	Words	Dale- Chall	Flesch	FOG	Powers*	SMOG	FOR- CAST	Frye	Spache	Adj Res	Pred Res
DORF Passages											
My Dress-Up Box	209	4.8	2.3	4.0	4.1	6.0	8.2	2.9	2.4	-3.4	-8.1
Mornings at Our House	252	4.6	1.9	4.8	4.0	6.2	7.9	2.5	2.1	-4.6	-4.4
Cleaning Your Plate	229	4.7	2.2	4.5	4.1	6.1	7.8	2.9	2.3	-7.3	-6.9
Sleeping Over	253	4.6	2.3	4.5	4.1	5.8	8.1	2.9	2.3	-1.6	-6.9
Colors of the Rainbow	214	4.4	2.1	4.1	4.1	5.7	8.4	2.7	2.5	-6.7	-9.4
I Want to Fly in Space	251	4.3	1.6	4.4	3.8	5.1	7.0	2.5	2.3	-9.1	-6.9
If I Had a Robot	233	5.2	2.0	4.5	4.0	5.7	7.7	2.6	2.4	0.6	-8.1
Mom's New Job	240	4.2	2.5	5.7	4.1	6.7	7.7	3.1	2.4	-5.2	-8.1
Our Camping Trip	226	4.4	2.0	7.8	4.1	6.6	8.0	2.6	2.1	-6.9	-4.4
DORF Average		4.6	2.1	4.6	4.0	6.0	7.9	2.7	2.3	-4.9	-7.0
TORF Passages											
Getting Reading To Go	212	4.2	2.3	5.6	4.0	6.9	7.1	2.9	2.2	-2.2	-5.6
The Race	202	5.6	3.4	5.3	4.2	5.3	8.0	3.9	2.8	-5.8	-13.1
King	175	4.1	1.1	3.3	3.6	4.2	7.0	1.9	1.8	-0.7	-0.7
Tap, Tap, Tap	179	4.1	0.7	2.9	3.6	4.4	7.4	1.5	1.8	3.5	-0.7
A Walk in the Woods	198	4.6	2.9	6.0	4.0	6.4	6.8	3.7	2.6	5.1	-10.6
TORF Average		4.5	2.1	4.6	3.9	5.4	7.3	2.8	2.2	0.0	-6.1

Note. Adj. Res. is adjusted mean residual for prior version of DORF and current version of TORF. Pred. Res. is predicted adjusted mean residual of current version based on Spache readability (Pred Res = 21.719 - 12.431 * Spache).

Reliability of DORF

The reliability of the selected DORF passages was examined by computing an inter-correlation matrix among the 9 DORF passages. The 9 passages provided 36 alternate-form reliability coefficients. The passages were administered between 1 and 5 weeks apart, and the time between administrations varied as a function of the random order of administration. The 36 alternate form reliability coefficients are reported in Table 2. The median alternate-form reliability coefficient for the DORF passages was .94. The median alternate-form reliability coefficient for the 10 inter-correlations among the TORF passages was .95.

Table 2

Alternate-Form Reliability of selected DORF Passages

DORF Passage	DORF Passage							
	1.	2.	3.	4.	5.	6.	7.	8.
1. Cleaning your plate								
2. My dress-up box	.91							
3. Mornings at our house	.96	.91						
4. I Want to Fly in Space	.95	.91	.95					
5. If I had a robot	.94	.89	.96	.94				
6. Colors of the rainbow	.94	.91	.95	.95	.94			
7. Sleeping over	.96	.91	.96	.95	.94	.95		
8. Our camping trip	.93	.90	.95	.94	.95	.93	.94	
9. Mom's new job	.94	.91	.96	.94	.93	.94	.94	.94

Note. All correlations were significant, $p < .01$.

Concurrent Validity of DORF with TORF

The concurrent validity of the DORF passages with respect to the TORF was examined by computing the correlation between each DORF passage and each TORF passage. The correlations between each of the selected DORF passages and the selected comparison TORF

passages are presented in Table 3. Correlations between DORF and TORF passages ranged from .92 to .96 and the median correlation was .95. The median inter-correlation of the TORF passages with each other also was .95.

Table 3

Concurrent, Criterion-Related Validity of DORF Passages with TORF Passages

DORF Passage	TORF Passage				
	A walk in the woods	Tap, tap, tap	Getting ready to go to the lake	The race	King
Cleaning your plate	.94	.95	.95	.95	.95
My dress-up box	.93	.92	.92	.93	.91
Mornings at our house	.94	.96	.95	.96	.96
I Want to Fly in Space	.94	.94	.95	.96	.95
If I had a robot	.93	.94	.94	.95	.93
Colors of the rainbow	.94	.94	.94	.95	.96
Sleeping over	.95	.94	.95	.95	.95
Our camping trip	.93	.93	.95	.95	.94
Mom's new job	.94	.94	.95	.95	.95

Note. Correlations based on sample size of 130 to 133. All correlations significant at $p < .01$.

In summary, the DORF passages

References

Children's Educational Services. (1987). Test of Reading Fluency. Minneapolis, MN: Author.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs. Chicago: Scientific Software International.

Micro Power & Light Co. (2000a). Readability Calculations According to Dale-Chall, Frye, et al. (Version 3.7e). Dallas, Tx: Author.

Micro Power & Light Co. (2000b). Readability Calculations According to the Spache Formula (Version 3.7s). Dallas, Tx: Author.