What's PSF Got To Do With It?

A Look At The Contribution of DIBELS® Phoneme Segmentation Fluency to First Grade Reading Outcomes

Kelly A. Powell-Smith, Ph.D. & Kelli D. Cummings, Ph.D. / Dynamic Measurement Group

Introduction & Rationale

The role of phonological awareness and, more specifically, phonemic awareness has been explored in numerous studies predicting later reading outcomes. In addition, much attention has been given to the role of rapid naming (RAN). In particular, the idea that children who are low in both RAN and PA experience the most severe reading difficulties has been investigated (see Wolf & Bowers, 1999).

Overall, research supports the important role that PA has in the development of reading skills (See Allor, 2002 for a review of several early studies; see also Anthony et al., 2006; Cardoso-Martins & Pennington, 2004; Compton, 2003; Kirby, Parrila, & Pfeiffer, 2003; Parrila, Kirby, & McQuarrie, 2004; Rouse & Fantuzzo, 2006; Schatschneider, Carlson, Francis, & Foorman, 2002; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Speece, Mills, Ritchey, & Hillman, 2003), with a few contrary findings (e.g., Evans, Bell, Shaw, Moretti, & Page, 2006; Morris, Bloodgood & Perney, 2003; Speece & Ritchey, 2005).

Despite this research, two issues have received little attention: (1) the use of fluency-based measures to index PA, and (2) confounded variability/shared variance between predictors. Researchers typically administer multiple measures of PA, yet few studies have explored the utility of fluency-based assessments. Examining fluency-based measures is critical because fluency with reading sub-skills is an additional important indicator of reading success (Conrad & Levy, 2006; Good, Simmons, & Kamme'nui, 2001; Logan, 1997a; 1997b) and fluency-based measures (i.e, DIBELS) now are commonly used to screen and monitor students' progress in learning to read. Regarding the second issue of shared variance between predictors, some studies have noted this finding (e.g., Cardoso-Martins & Pennington, 2004; Schatschneider et al., 2002) and others have explored the relationship directly in terms of reciprocal causation (e.g., Compton, 2003). Still, most research has focused on determining unique variance and thus, shared variance largely has been ignored.

This study explores the relation of DIBELS Phoneme Segmentation Fluency (PSF) to reading outcomes. Although research demonstrates linkages between each DIBELS measure (i.e., PSF predicts Nonsense Word Fluency (NWF), NWF predicts Oral Reading Fluency (ORF) and ORF predicts outcomes on a statewide test; see Good et al., 2001), we found only one recently published study (Rouse & Fantuzzo, 2006) that examined a more direct link of PSF to reading outcomes. This study also explores PSF's shared contribution (with NWF and ORF) to reading outcomes.

Research Questions

1. What is the explanatory value added from screening and monitoring with PSF in first grade when predicting ORF outcomes?

2. What is the explanatory value added from screening and monitoring with PSF in first grade when predicting SAT-10 outcomes?

3. Does the effect of PSF on later reading outcomes change depending on LNF risk status?

4. What percent of explained variance in reading outcomes is shared with other measures of early reading?

Method

Unless otherwise mentioned, risk categories are based on student scores from the LNF measure in the fall of first grade.

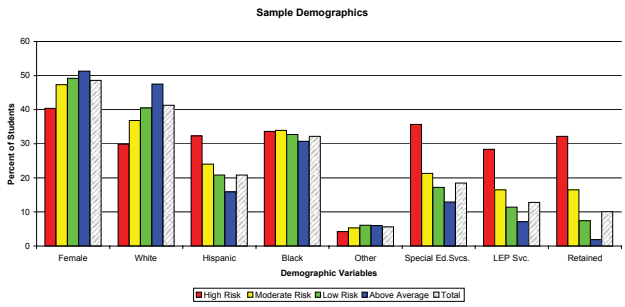


Figure 1. Demographic Characteristics of the Sample

Note. The category of "Other" race includes students who identified as belonging to Asian or Pacific Islander, American Indian or Alaska Native, or Multiracial/Multiethnic groups. The category of "Special Ed. Svcs." *excludes* students who were receiving special education under the category of gifted. "LEP Svc." refers to the percent of students who were identified as Limited English Proficient and/or were receiving services for Limited English Proficiency at the time of data collection. "Retained" students were held back at least one year.

Participants First grade students in the first year (2003-2004) of Reading First implementation in Florida participated in this study. The sample consisted of 27,813 first grade students from 321 schools. Only students with complete data on both outcome measures (ORF-administered at Time 4, and SAT-10) were included in the study. In addition, students' scores were deleted if they fell outside the range of the "refuse levels" from the DIBELS® Data System. Demographic and descriptive information on the participants is shown in Figure 1 and Table 1.

Measures The four measures used in this study were: (a) Phoneme Segmentation Fluency (PSF; Kaminski & Good, 1996). PSF assesses a student's ability to segment three- and four-phoneme words into their individual phonemes fluently. The number of correctly produced phonemes in one minute is the score. (b) Nonsense Word Fluency (NWF; Good & Kaminski, 2002). The NWF task is a measure of the alphabetic principle—including both letter-sound correspondence and the abil-

ity to blend letters into words. The student is presented with randomly ordered VC and CVC nonsense words and asked to produce verbally the individual letter sound of each letter or read the whole nonsense word. The final score is the number of letter-sounds produced correctly in one minute. (c) Oral Reading Fluency (ORF). The DIBELS® ORF (DORF, Good, Kaminski, & Dill, 2002) are a set of generic reading passages. Student performance is measured by having students read novel connected text aloud for one minute. The number of words read correct within one-minute is the score. (d) Stanford Achievement Test-10 (SAT-10) Reading Comprehension. This is a published norm-referenced test designed to assess reading comprehension. Students are required to read text passages and then answer questions.

Procedures This study was a secondary analysis of existing data obtained from the Florida Center for Reading Research's (FCRR) Progress Monitoring and Reporting Network (PMRN). The data for this study were obtained following FCRR PMRN data request procedures.

Because of the large sample size in this study, we adopted Good, Baker, & Peyton's (in press) application of Cohen's (1988) effect size criteria for evaluating magnitude of effects. Thus, an effect that explains 10% or more of the variance is considered a medium effect and of sufficient magnitude that it is worthy of our attention.

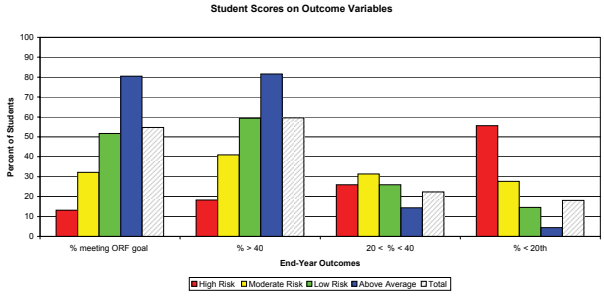


Figure 2. Student Scores on ORF4 and SAT-10

Results indicate that LNF may function less well as a predictor of end-of-first grade outcomes when administered at the beginning of the year. Nearly half of the students considered low risk on this measure performed below the benchmark goal on ORF, and fifty-nine percent scored at grade level on the SAT-10.

Table 1. Means and Standard Deviations for DIBELS® measures across a school year

	LNF	PSF				NWF				ORF		
Group		**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**
High Risk	15.33	15.81	28.24	37.69	38.35	8.25	22.00	27.68	36.98	2.14	5.62	11.92
(*n* = 3336)	(7.15)	(15.43)	(17.62)	(18.03)	(15.99)	(9.96)	(15.95)	(18.86)	(22.69)	(5.44)	(8.00)	(13.26)
Mod. Risk	31.33	25.44	36.86	45.48	44.26	18.02	33.99	39.44	48.94	5.61	10.50	21.06
(*n* = 6095)	(3.47)	(16.62)	(15.24)	(14.08)	(12.98)	(12.11)	(14.68)	(17.30)	(21.18)	(8.32)	(11.11)	(16.40)
Low Risk	41.28	30.98	40.48	48.03	45.97	26.25	41.58	47.56	57.26	10.60	16.57	30.16
(*n* = 6859)	(2.86)	(16.31)	(13.83)	(12.69)	(12.08)	(14.53)	(16.85)	(20.31)	(24.76)	(13.19)	(16.32)	(21.22)
Above Av.	57.74	37.18	43.96	50.32	47.96	40.57	55.94	64.01	74.87	23.14	32.14	50.11
(*n* = 11523)	(9.08)	(16.06)	(13.02)	(11.91)	(11.85)	(21.17)	(24.73)	(28.37)	(32.78)	(23.43)	(26.71)	(29.84)
Total	42.81	30.51	39.66	47.18	45.51	28.22	43.52	50.22	60.30	13.69	20.38	34.25
(*n* = 27813)	(16.21)	(17.65)	(15.19)	(14.02)	(13.07)	(20.41)	(23.27)	(26.65)	(30.66)	(18.94)	(22.46)	(27.68)

Note. All students whose data are entered in Florida's PMRN database are assessed four times per year, corresponding to the data listed above in columns 1–4.

In general, student scores improved over the course of the school year. However scores on PSF tended to remain constant, or to slightly decrease, between assessment periods three and four. All scores were significantly ($p < .01$) different from one another due to the large sample size. Groups were significantly different, with large effect sizes ($\eta 2$ greater than .25), on NWF measures at times 1 and 2, ORF measures at times 3 and 4, and on the SAT-10.

Results

Table 2. Correlation Matrix for DIBELS® measures Across Time, With SAT-10

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1. LNF	____	.42	.34	.28	.24	.64	.58	.56	.51	.52	.54	.59	.60	.55
2. PSF1		____	.57	.45	.38	.46	.39	.34	.32	.29	.29	.33	.33	.40
3. PSF2			____	.58	.48	.34	.38	.31	.30	.18	.22	.26	.28	.33
4. PSF3				____	.56	.26	.29	.32	.27	.13	.15	.21	.23	.27
5. PSF4					____	.23	.25	.25	.30	.11	.12	.16	.20	.23
6. NWF1						____	.76	.70	.64	.73	.73	.73	.70	.60
7. NWF2							____	.77	.71	.67	.72	.73	.71	.59
8. NWF3								____	.79	.63	.69	.76	.74	.61
9. NWF4									____	.57	.63	.71	.74	.60
10. ORF1										____	.93	.85	.76	.58
11. ORF2											____	.91	.83	.64
12. ORF3												____	.93	.75
13. ORF4													____	.79
14. SAT10														____

Correlation coefficients between measures of Oral Reading Fluency and SAT-10 outcomes are similar to others found between measures of Oral Reading Fluency and Statewide Achievement Tests (Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). Correlations between NWF and SAT-10 are also high, and relatively stable across the course of the first grade year. Correlations between PSF and SAT-10 decrease over the course of the year, highlighting a possible threshold effect with the measure beyond which scores function less well as predictors.

Variables were entered sequentially into hierarchical regression models to predict end of year reading outcomes as measured by the SAT-10 and DIBELS®-Oral Reading Fluency (DORF). The rationale for the order of the variables was due to (a) the notion of a developmental progression of early reading skills and (b) the nature of all predictors as malleable, or amenable to intervention.

Table 3a. Variance in End-of-Year Outcomes Explained for Each Risk Category

		Sequential percent of variance in SAT-10 explained within risk category		
Group	**Regression Equation**	**PSF**	**NWF given PSF**	**ORF given NWF & PSF**
High Risk	$\hat{Y} = 487 + .26PSF + 1.12NWF + 1.44ORF$	.07	.22	.24
Moderate Risk	$\hat{Y} = 500 + .25PSF + .77NWF + 1.65ORF$	.07	.20	.29
Low Risk	$\hat{Y} = 510 + .28PSF + .67NWF + 1.17ORF$	.06	.23	.31
Above Average	$\hat{Y} = 535 + .37PSF + .23NWF + .80ORF$	.05	.21	.31
Overall	$\hat{Y} = 505 + .48PSF + .70NWF + .85ORF$	.16	.38	.43

		Sequential percent of variance in ORF4 explained within risk category	
Group	**Regression Equation**	**PSF**	**NWF given PSF**
High Risk	$\hat{Y} = 11 + .01PSF + 1.14NWF$	.06	.33
Moderate Risk	$\hat{Y} = 18 - .02PSF + .91NWF$	.03	.23
Low Risk	$\hat{Y} = 24 - .07PSF + .92NWF$	.02	.26
Above Average	$\hat{Y} = 33 - .05PSF + .91NWF$	.02	.36
Overall	$\hat{Y} = 18 + .02PSF + 1.09NWF$	.11	.49

With our a priori criteria for determining importance of effect sizes (R2> .10), both PSF and the other predictors in the model explained important components of later reading outcomes. Using a battery of approxi-

mately six minutes of testing in the beginning of the year, we were able to explain over 43% of the variability in SAT-10 outcomes. Taken together, the PSF and NWF measures in the fall, explained nearly half of the variability in oral reading at the end of first grade.

In order to determine the shared contribution of all predictors, in addition to their unique variance components, we reversed their order in two regression equations. The results are summarized in Table 4.

Table 4. Summary of Two Hierarchical Regression Analyses for Variables Predicting SAT-10 (N = 27813)

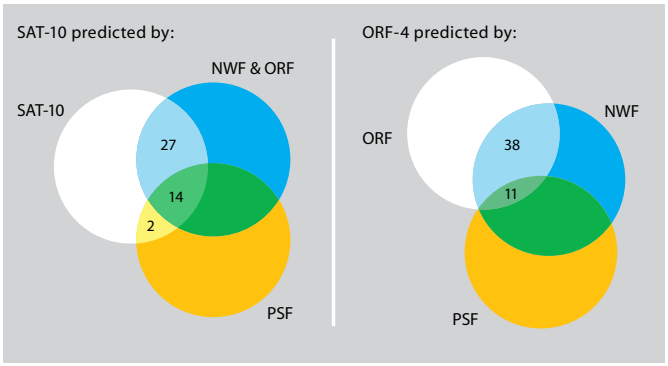
	Regression and Steps	B	SE B	β	R^2	ΔR^2
A	1. PSF	1.11	.02	.40	.16	
	2. NWF	.70	.02	.29	.43	.27
	ORF	.85	.02	.33		
B	1. NWF	.92	.02	.38	.41	
	ORF	.80	.02	.31		
	2. PSF	.48	.01	.17	.43	.02

Note. All beta weights were significantly different from zero (p <.00).

Summary of Two Hierarchical Regression Analyses for Variables Predicting ORF4 (N = 27813)

	Regression and Steps	B	SE B	β	R^2	ΔR^2
A	1. PSF	.61	.01	.33	.11	
	2. NWF	1.09	.01	.69	.49	.38
B	1. NWF	1.10	.01	.70	.49	
	2. PSF	.02	.01	.01	.49	.00

The Venn Diagrams below indicate the degree of shared and unique variance for the various predictors of reading outcomes. Note that, when predicting SAT-10 outcomes, NWF and ORF were combined.



Discussion

Implications

The finding that the amount of unique variance accounted for by PSF was small in the presence of the other predictors does not necessarily mean that it is unimportant for the development of reading. On the contrary, we believe not only is this measure is an important predictor but that it may have a synergistic relationship with other early literacy skills. It has already been said that PA "may lose its predictive power when variables more closely related to reading are included" (Speece & Ritchey, 2005, p. 397), but this statement does not address what is shared between PA and these variables.

There are various ways one might talk and think about shared variance. The first is that the measures assess the same constructs. Another is reciprocal causation (bidirectional relationship) (see Compton, 2003; Perfetti, Beck, Bell & Hughes, 1987; Wagner, Torgeson, & Rashotte, 1994). A third possibility is that the shared variance is related to developmental progression where later, or more advanced, skills may incorporate earlier and less advanced skills. Finally, one might conceive of shared variance as representing synergy.

We resonate most with the synergy explanation because it implies that in order to have the biggest impact on outcomes, one must teach, change, and integrate each of the represented skill areas. For example, one must teach, change and integrate the skills indexed by PSF and those indexed by NWF. Doing so, not only would hypothetically produce the gains consistent with the unique variance attributable to each, but also the gains attributed to what they share. Importantly, we believe the idea of synergy is consistent with one of the six major principles of effective instructional tools known as "strategic integration" (see Kamme'enui, Carnine, Dixon, Simmons, & Coyne, 2002). Strategic integration requires "the careful systematic combining of essential information in ways that result in new and more complex knowledge" (Kamme'enui et al., p. 13). Further, we believe this approach is consistent with a multicausal system explanation for reading development where "different processes interact to determine outcomes" (Hulme, Snowling, Caravolas, & Carroll, 2005, p. 352). It is the nature of this interaction that needs greater exploration.

Limitations

The primary limitation of this study results from using an existing data set. Scoring inaccuracies would lead to spurious results and some measures (i.e., PSF) may be more challenging to administer than others (i.e., ORF). We believe that this concern highlights the need for DIBELS® users and researchers to (a) use an accuracy of assessment checklist and (b) retest a sample of students to check for reliability of scores.

Considerations for Future Research

Future research should: (1) replicate the analysis with a kindergarten sample and examine the role of PSF longitudinally across Kindergarten and into first grade, and (2) attempt to further unpack the nature of the shared variance between early literacy measures. Research incorporating the use of latent variables may determine the extent to which shared variance is attributable to a common index between predictor measures (i.e., "fluency" alone) or multiple skills that should be incorporated into instruction.