

# Early Classification of Reading Performance in Children Identified or At Risk for Emotional and Behavioral Disorders: A Discriminant Analysis Using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

Jorge E. Gonzalez, Kimberly J. Vannest, and Robert Reid

**Abstract:** This study evaluated the ability of the kindergarten and first grade Dynamic Indicators of Basic Early Literacy Skills (DIBELS), measures of early literacy development, to discriminate among low average, average, and above average students considered at risk emotional and behavioral disorders (EBD) on the Total Reading cluster of the Woodcock Reading Mastery Tests-Revised (WRMT-R). The DIBELS consisted of two measures of phonological awareness, one measure of alphabet knowledge, one measure of the alphabetic principle, and one measure of oral reading fluency with connected text. Results indicated that first grade DIBELS differentiated among reading groups and classification accuracy was statistically better than chance. With the exception of alphabet knowledge, DIBELS did not significantly differentiate among the fall kindergarten groups. Oral reading fluency and alphabet knowledge had the greatest discriminating power for first graders. These findings extended the usefulness of the first grade DIBELS to populations other than general education students. Implications for the use and application of DIBELS to non-general education populations are discussed along with caveats for kindergarten discriminant power of the DIBELS.

## Introduction

No one would argue that early intervention is essential for preventing or mitigating the impact of emotional and behavioral disabilities (EBD) on academic performance. This may be especially crucial in reading. The fundamental nature of reading ostensibly serves as the fulcrum for a majority of other learning demands. There currently exists a need to pay more focused attention to early identification and prediction of correlated behavior and reading difficulties (Trout, Nordness, Pierce, & Epstein, 2003).

Academic failure is pervasive for students with EBD. For these students, EBD will persist over time often disrupting social, academic, and community functioning (Kutash, Duchnowski & Friedman, 2005). Approximately 38% of students identified as EBD have been retained by the time they reach secondary school (Wagner, Kutash, Duchnowski, Epstein & Sumi, 2005) with most 1.5 to 3 grade levels below same age peers (Coutinho, 1986; Trout et al., 2003). While exact numbers vary, approximately 60% of elementary/middle school children with EBD perform in the bottom quartile on reading measures with 85% making up the bottom two quartiles (Wagner et al., 2005). Moreover, students identified with EBD are consistently found to have the highest school dropout incidence rates in children and youth identified with disabilities (Reschly & Christenson, 2006).

Although there is still much to be done on identifying and predicting which children are at risk for reading failure, a source of urgency is the relationship between reading failure and concomitant development of emotional and behavioral problems. Academic performance has consistently been shown to be inversely related to problem behavior beginning early in a child's schooling (Brier, 1995; McEvoy & Welker, 2000). Students with poor reading skills are more likely to experience negative behavioral and or antisocial outcomes in the future (Good, Gruba, & Kaminski, 2001; Good, Simmons, & Smith, 1998; McEvoy & Welker, 2000). The early identification and prevention of academic deficits, particularly in reading, may assuage and ultimately diminish the development of behavioral problems.

Results of four systematic research reviews synthesizing dozens of studies show that early reading failure is correlated with the onset, persistence, and seriousness of emotional and behavioral problems independent of major sociological variables such as socioeconomic status (Gottfredson, 1981; Hawkins & Lishner, 1987; McEvoy & Welker, 2000; Silberberg & Silberberg, 1971). These reviews show that poor academic achievement and/or academic survival skills often coincide with behavioral difficulties (Cullinan & Epstein, 2001; Cullinan, Evans, Epstein, & Ryser, 2003). Depending on the criteria for identification of EBD, studies demonstrate a 6% to 42% comorbidity of emotional disturbance involving academic difficulties.

Given the relationship among emotional, behavioral, and reading problems, one could reasonably assume that there would be a preponderance of research on the early detection of reading difficulties in students with EBD or vice versa and on interventions to address these problems. To the contrary, there is no preponderance of research in this important area (Coleman & Vaughn, 2000; Lane, 2004; Mooney, Epstein, Reid & Nelson, 2003; Ruhl & Burlinghoff, 1992; Trout et al., 2003; Wagner et al., 2005).

If researchers and educators are to improve academic and behavioral outcomes for low-achieving, beginning readers, measures of proficiency in critical early reading skill areas are essential. Conventional standardized student reading achievement data, unfortunately, do not yield sufficient information on these skills to differentiate successful from less successful readers. Traditional reading measures only provide summative information, infrequent measurement points, omit student progress monitoring, and have little instructional utility (Good et al., 2001; Good et al., 1998). Given accumulating evidence that reading success is causally influenced by ease with critical early reading skills, valid, formative, and reliable assessment tools are needed to determine performance on these skills before students begin to learn to read (Elliot, Lee, & Tollefson, 2001; Kaminski & Good, 1996).

One set of measures useful in monitoring progress on early reading skills are the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Good et al., 2001). The DIBELS are a continuum of fluency-based measures that assess facility in pre-reading and early reading skills, are highly correlated to later reading competence, and are aligned with scientifically-based reading research (Good et al., 1998; Hintze, Ryan, & Stoner, 2002). These skills include initial sound fluency, letter naming fluency, phonemic segmentation, nonsense word reading, and oral reading fluency. The DIBELS permit: (a) early identification of students with reading difficulties, (b) formative evaluation of instructional effectiveness in pivotal skills, and (c) determination of reading development (Kaminski & Good, 1996).

The reliability and validity of the DIBELS for heterogeneous samples of general education elementary students has been investigated in several reports and studies (Buck & Torgesen, 2003; Elliot et al., 2001; Hintze et al., 2002; Kaminski & Good, 1996; Shaw & Shaw, 2002). Alternate form reliability estimates for the DIBELS measures are generally adequate, ranging from .72 for *Initial Sound Fluency* (ISF) to .93 for *Letter Naming Fluency* (LNF). Elliott (2001) found test-retest reliability estimates of .90, .83, and .85 for LNF, Sound Naming Fluency (adaptation of INF), and Phonemic Segmentation Ability (adaptation of PSF), respectively. Concurrent validity estimates for the DIBELS measures range from a low of .09 for PSF (e.g., *Comprehensive Test of Phonological Processing*) (Wagner, Torgesen, & Rashotte, 1999) to a high of .85 for LNF (with teacher rating scale). Predictive validity estimates with well known measures, like the Woodcock-Johnson Psycho-educational Battery (Woodcock, Johnson, Mather, McGrew, & Werder, 1991), are adequate, ranging from .45 for *Oral Reading Fluency* (ORF) to .71 for *Letter Naming Fluency* (LNF).

The reliability and validity of the DIBELS has been well established among general education students; the ability to identify students who are at risk of failure is a separate issue. Only one study has evaluated the predictive validity of the DIBELS with this population. Hintze et al., (2002) studied a sample of 86 general education kindergartners. Results indicated that the combination of ISF, PSF, and LNF were ac-

curate in predicting membership in a poor reading group. Prediction accuracy for the high reading group was somewhat mixed. In other words, the DIBELS measures lead to a very high percentage of true positives (i.e., children correctly identified as performing poorly on the *Comprehensive Test of Phonological Awareness*). While the results of this study are encouraging, their generalizability is limited. We do not know if the DIBELS will be an accurate predictor of reading performance for children at risk of behavior problems or disorders.

If educators and researchers are to use the DIBELS to identify at-risk readers, then it is important to substantiate their validity across student populations other than general education elementary students. This study contributes to the research base by examining the usefulness of the DIBELS in differentiating among low average, average, and above average students identified at risk of emotional and behavioral disturbance based on performance on the Woodcock-Johnson Reading Mastery test.

## Methods

### Participants and Setting

Participants were 145 students (67 kindergartners, 78 first graders) identified as at risk of emotional and behavioral disorders (EBD) (see Table 1 for demographic characteristics of sample participants). The students were drawn from seven elementary schools in a medium-sized urban school district in the Midwest. The schools selected were part of a school district with a total minority enrollment of 15%. Districtwide, 27% of students receive free or reduced lunch and 15% receive special education services with over 5% of children in English Language Learner (ELL) programs. The students were drawn from an initial pool of 322 students invited to participate in research on the effects of primary (e.g., universal programs like mentoring programs); secondary (targeted programs like counseling and social skills training); and tertiary (intensive programs like behavior programs, related services) levels of intervention for students identified at risk of EBD.

Table 1  
Demographic Characteristics of Sample Participants

Characteristic	Kindergarten (n = 67)	First Grade (n = 78)
Age		
Mean	5.6	6.2
Standard Deviation	.38	.40
Gender		
Male	56 (82%)	51 (66%)
Female	12 (18%)	26 (34%)
Ethnicity		
White	44 (64%)	54 (69%)
African American	15 (22%)	13 (17%)
Hispanic	7 (11%)	8 (10%)
Other	2 (03%)	3 (04%)

Note: Numbers in parentheses are percentages.

Students were identified as “at risk” using the *Systematic Screening for Behavior Disorders* (SSBD) program (Walker & Severson, 1990). The SSBD consists of three “gates” that provide progressively more intensive levels of screening; however, only the first two steps were necessary for subject identification in this study. In the spring and fall of 2002, all kindergarten and first grade students in the seven participating elementary schools were screened. In Stage I, teachers identified the five students who demonstrated the most problematic externalizing characteristics (e.g., acting out, aggressiveness, bullying) and five students who exhibited the most internalizing characteristics (e.g., anxiousness, depressed mood, sadness). In Stage II, teachers completed three measures for each of the 10 students identified in Stage I, namely: (1) the Critical Events Index, (2) the Adaptive Behavior Scale, and (3) the Maladaptive Behavior Scale. For the Critical Events Index, the teacher indicated the occurrence or nonoccurrence of 33 externalizing and internalizing problem behaviors over the course of the previous six months. The Adaptive Behavior Scale (12 items) indicated the presence or lack of prosocial behavior (e.g., “follows established classroom rules”). The Maladaptive Behavior Scale (11 items) indicated the presence or lack of antisocial behavior (e.g., “refuses to participate in games and activities with other children at recess”). Those students who exceeded the normative criteria on one or more of the measures in Stage II were considered as at risk of EBD and included in the study.

## Instruments

*Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*. All DIBELS are reported as a frequency per minute rate total score. DIBELS data collected in this study included *Initial Sound Fluency* (ISF: kindergarten); *Letter Naming Fluency* (LNF: kindergarten, first grade); *Phoneme Segmentation Fluency* (PSF: kindergarten, first grade); *Nonsense Word Fluency* (NWF: kindergarten, first grade); and *Oral Reading Fluency* (ORF: first grade).

*Initial Sound Fluency* (ISF) is a measure designed to assess a student’s facility in recognizing and producing the initial sound from an orally presented word. For example, using four pictures, the examiner asks “This is a sink, cat, gloves, and a hat. Which picture begins with /s/?” Each probe consists of 12 items. The ISF task takes about three minutes to administer and possesses an alternate form reliability of .72 (Hintze et al., 2002).

*Letter Naming Fluency* (LNF) assesses rapid letter naming ability. Students are presented with a page of random upper- and lowercase letters and are asked to name as many letters as they can. The student is allowed one minute to produce as many letter names as he/she can with the score being the number of letters named correctly in one minute. Alternate form reliability for the LNF is .93 (Hintze et al., 2002).

*Phoneme Segmentation Fluency* (PSF) assesses a student’s ability to segment three- and four-phoneme words into their individual phonemes fluently. A phoneme is the smallest unit of sound which is significant in a language. It requires the student to verbally produce the individual phonemes for each word. The PSF measure takes about two minutes to administer and has over 20 alternate forms for progress monitoring.

*Nonsense Word Fluency* (NWF) assesses the alphabetic principle. This includes letter-sound correspondence along with the blending

of letters into words (Kaminski & Good, 1996). The student is presented with random VC (Vowel Consonant) and CVC (Consonant, Vowel, Consonant) nonsense words (e.g., sig, rav, ov) and is asked to either produce letter sounds in isolation or to orally produce the whole nonsense word. The score is the total number of letter-sounds produced correctly at the end of one minute.

*Oral Reading Fluency* (ORF) assesses reading with accuracy and fluency. Students read aloud a grade-appropriate passage for one minute. Omissions, substitutions, and hesitations greater than three seconds on words in the passage are considered errors. The oral reading fluency rate is determined by the number of words read correctly within one minute.

*Woodcock Reading Mastery Test-Revised*. The Woodcock Reading Mastery Test-Revised (WRMT-R) (Woodcock, 1998) is an individually administered, norm-referenced, standardized measure used to assess students’ beginning reading skills. The test mean is 100; the standard deviation is 15. The WRMT-R provides a total score for overall reading (i.e., Total Reading cluster which serves as a broad measure of global reading ability) and six subtest scores. Subtests (and coefficient alphas) are: Word Identification (alpha = .97) measures ability to pronounce words in isolation; Word Attack (alpha = .91) measures ability to use phonic and structural analysis to pronounce nonsense words; Word Comprehension (alpha = .91) measures vocabulary skills using Antonyms, Synonyms, and Analogy Completion tasks; Passage Comprehension (alpha = .92) measures ability to supply a missing word from a brief passage; Reading Comprehension (alpha = .95) combines Word Comprehension and Passage Comprehension; and Basic Skills Cluster (alpha = .97) combines Word Identification and Word Attack.

## Data Collection Procedures

Measures were collected during the fall of 2002. All measures were administered by data collectors who had received 20 hours of formal training in the administration and scoring of each measure. All data collectors were required to demonstrate mastery (i.e., 90 – 100% inter-rater agreement) prior to testing. In one session, all students were tested individually with the DIBELS and the WRMT-R. Testing sessions took approximately 20 and 45 minutes, respectively. According to DIBELS instructions, kindergartners were tested on ISF, LNF, NSW, and PSF; and first graders were tested on LNF, NSW, PSF, and ORF. Each student was allowed a break between measures. In the event an administration was not completed due to unforeseen circumstances (e.g., fire drill), a second administration was scheduled no more than one week later.

## Data Analysis

Two separate hierarchical discriminant analyses were conducted, one each for first grade and kindergarten, respectively. There are two kinds of conventional discriminant analysis studies—descriptive discriminant analysis (DDA) and predictive discriminant analysis (PDA). The primary objective of DDA is to identify attributes or variables that best discriminate members of two or more groups. PDA is used primarily to predict group membership in mutually exclusive groups of two or more (Duarte-Silva & Stam, 2004). Unlike conventional discriminant analysis in which a set of variables is used to predict group membership, hierarchical discriminant analysis permits the effect of

each single variable to be studied uniquely. In this article, separate analyses were conducted for first grade and kindergarten because kindergarten students are not measured on ORF, and first graders are not given ISF. Students were divided into three groups based upon the total cluster WRMT-R Reading scores: low average, average, and high average. For kindergarten, less than and equal to 89 was low average, 90 to 103 was average, and 104 or higher was above average. For first graders, less than and equal to 86 was low average, 87 to 98 was average, and 99 or higher was above average. In each case, group membership (i.e., low, average, and above average) was the dependent variable. For kindergarten and first grade, the independent variables were ISF, LNF and PSF and LNF, PSF, NWF and ORF. A preliminary analysis was conducted to examine the distribution of variables and their correlation matrix. Results showed that kindergarten NWF and first grade ORF and NWF were highly skewed. To normalize the skewness, square root transformations were employed. To control for the possible violation of homogeneity of variance-covariance, separate (as opposed to pooled) variance-covariance matrices in the classification were used for first grade since the Box's test indicated that the assumption had been violated.

Because there were no *a priori* assumptions about group membership, prior probabilities were set at equal group membership. The discriminatory power of the classification matrix in comparison to a chance model was tested using Press's Q statistic (Hair, Anderson, Tatham, & Black, 1998). If Press's Q statistic exceeds the chi-square critical value, then the classification matrix is considered statistically better than chance. Classification accuracy for low average, average, and above average groups was tested using the proportional chance criterion (Hair et al., 1998). These criteria provide a test of how accurately each group could be classified in relation to the total sample. In this study, hierarchical discriminant analysis was employed following the step down theory (Duarte-Silva & Stam, 2004; Roy & Bargman, 1958) with separate analyses for kindergarten and first grade students, respectively. Descriptive discriminant analysis (DDA) was applied to the set of DIBELS which most contributed to predicting WRMT-R scores. ANCOVA was subsequently employed on each additional predictor one at a time to test whether it had a significant effect on WRMT-R score group membership controlling for previous predictors already included in the model as covariates. The predictor was included in the model if it significantly influenced the WRMT-R group membership.

## Results

Table 2 contains descriptive statistics for kindergarten and first grade children on both the DIBELS and the WRMT-R Total Score. Kindergarten WRMT-R Total standard scores assumed a bimodal distribution with first grade WRMT-R assuming a normal distribution. Table 3 shows the correlation matrices for both kindergarten and first grade. For kindergarten, only LNF was significantly correlated with the WRMT-R. In contrast, all the first grade DIBELS were significantly correlated with WRMT-R standard scores.

Table 2

Descriptive Statistics for Kindergarten and First Grade Samples

	Variable	Mean	Median	SD	Min.	Max
Kindergarten	ISF	12	10	7	0	32
	LNF	18	17	13	0	51
	NWF	4	0	8	0	37
	PSF	6	0	9	0	36
	WRMT-R	98	101	11	81	116
First Grade	LNF	35	34	18	3	84
	NWF	20	18	18	0	95
	PSF	20	19	16	0	61
	ORF	10	5	17	0	93
	WRMT-R	91	91	14	57	124

Note. ISF = Initial Sound Fluency; LNF = Letter Naming Fluency; NWF = Nonsense Word Fluency; PSF = Phoneme Segmentation Fluency; ORF = Oral Reading Fluency; WRMT-R = Reading Mastery Total Reading Cluster Standard Score.

Table 3

Intercorrelations Between DIBELS and WRMT-R (TR)

Variable	TRC	ISF	LNF	NWF	PSF	ORF
Kindergarten (n = 67)						
WRMT-R	-	.05	.33**	.22	.15	-
ISF		-	.21	.20	.46**	-
LNF			-	.51**	.29*	-
NWF				-	.60**	-
PSF					-	-
ORF						-
First Grade (n = 77)						
WRMT-R			.66**	.71**	.54**	.68**
LNF			-	.73**	.66**	.53**
NWF				-	.56**	.79**
PSF					-	.43**
ORF						-

Note. DIBELS = Dynamic Indicators of Basic Early Literacy Skills; Literacy Skills; WRMT-R = Woodcock-Johnson Reading Mastery-Revised Total Cluster Score; ISF = Initial Sound Fluency; LNF = Letter Naming Fluency; NWF = Nonsense Word Fluency; PSF = Phoneme Segmentation Fluency; ORF = Oral Reading Fluency.

\*Significant at 0.05.

\*\*Significant at .01.

In the hierarchical discriminant analysis of kindergarten students, DDA was employed on the model with LNF and ISF as predictors because they had the largest correlation with the discriminant function of WRMT-R. The results for the kindergarten group were marginally significant for Function 1 ( $\Lambda = .86$ ),  $\chi^2(4, N = 67) = 9.96, p = .041$ . Function 1 accounted for 82% of the explained variance. Standardized discriminant function coefficients for deriving discriminant function scores from standardized predictors were 1.01 for LNF and -.44 for ISF. Correlations (loadings) between LNF and ISF and the discriminant functions given in the structure matrix were .91 and -.20 for LNF and ISF respectively. The structure matrix of correlations between the predictors and the discriminant functions suggested that LNF was the best predictor for distinguishing between low, average, and above average students.

The effect of ISF, while not noteworthy, had a suppression effect and increased the effect size of the study. Therefore, both LNF and ISF remained in model. In the second step, ANCOVA was applied to PSF controlling for LNF and ISF as covariates. In the third step, ANCOVA was applied to NWF controlling for the other three predictors as covariates. Results showed that neither NWF or PSF had a significant effect on WRMT-R group membership. The final model had two DIBELS measures employed for the PDA of kindergarten students. Group means and standard deviation of LNF and ISF of three WRMT-R group levels are shown in Table 4. The use of the two DIBELS measures resulted in an overall classification accuracy of approximately 50% with 68.2% of low level, 50% of average level, and 32% of high level groups respectively as represented in Table 5. It illustrates that LNF and ISF are much more efficient in predicting low level reading performance ability students than high level kindergarten students. It reveals that DIBELS might not be good at predicting high WRMT-R level of students who were identified at risk of emotional and behavioral disorders.

Table 4

Group Means and Standard Deviation for Kindergarten Grade Readers

Group	Variable	Mean	Standard Deviation
1	Letter Naming Fluency (LNF)	11.8	10
	Initial Sound Fluency (ISF)	12	7.6
2	Letter Naming Fluency (LNF)	21	14
	Initial Sound Fluency (ISF)	9.8	5.7
3	Letter Naming Fluency (LNF)	20.5	13.4
	Initial Sound Fluency (ISF)	13	8
Total	Letter Naming Fluency (LNF)	17.8	13.1
	Initial Sound Fluency (ISF)	11.7	7.3

Table 5

Classification Results for Kindergarten Students

Reader Group	Predicted Reader Group Membership		
	(LA)	(A)	(HA)
Low Average (LA)	15 (68.2%)	5 (22.7%)	2 (9.1%)
Average(A)	6 (30%)	10 (50%)	4 (20%)
High Average (HA)	10 (40%)	7 (28%)	8 (32%)

Note. 49.3% of grouped cases correctly classified.

For first grade students, the DDA was applied on the ORF and LNF predictors because they had the strongest correlation with the discriminant function. The results for the first grade group were significant for Function 1 ( $\Lambda = .45$ ),  $\chi^2(4, N = 78) = 60.1, p \leq .001$ . Function 1 accounted for 96.3% of the explained variance. For first grade, standardized discriminant function coefficients for the discriminant function scores from standardized predictors were .53 for LNF and .67 for ORF. Correlations (loadings) between LNF and ORF and the discriminant functions given in the structure matrix were .88 and .27 for LNF and ORF, respectively. The structure matrix of correlations between the predictors and the discriminant functions suggested that both LNF and ORF were good predictors for distinguishing between low, average, and above average students.

Following the DDA, ANCOVA was applied to NWF, which was the third most important predictor of WRMT-R, using ORF and LNF as covariates. It was shown that NWF had no significant effect on the WRMT-R groups. Finally, ANCOVA was applied to PSF with ORF, LNF and NWF as covariates. Results showed that PSF still did not have a statistically significant affect on WRMT-R groups. The final model included only ORF and LNF as predictors in first grade and a PDA was applied to them. Table 6 shows the first grade ORF and LNF group means and standard deviation. The three WRMT-R levels increased with higher measured reading ability. Results show that 68% of original cases were correctly classified with 65.5% of low level, 68% of average level, and 71% of high level groups respectively and is shown in Table 7. Therefore, ORF and LNF were the most efficient in predicting high level reading ability and moderately predicted average and low level of reading for first grade students.

Table 6

Group Means and Standard Deviation for Grade Readers

Group	Variable	Mean	Standard Deviation
1	Letter Naming Fluency (LNF)	22	11.6
	Oral Reading Fluency (ORF)	1.2	.96
2	Letter Naming Fluency (LNF)	36.6	11.5
	Oral Reading Fluency (ORF)	2.1	1.1
3	Letter Naming Fluency (LNF)	50.3	17.9
	Oral Reading Fluency (ORF)	4.5	2.2
Total	Letter Naming Fluency (LNF)	35.4	18
	Oral Reading Fluency (ORF)	2.5	2

Table 7

Classification Results for First Grade Students

Reader Group	Predicted Reader Group Membership		
	(LA)	(A)	(HA)
Low Average (LA)	19 (65.5%)	9 (31%)	21 (3.4%)
Average (A)	5 (20%)	17 (68%)	3 (12%)
High Average (HA)	1 (4.2%)	6 (25%)	17 (70.8%)

Note. 67.9% of grouped cases correctly classified.

## Discussion

Results of the present study showed that the first grade DIBELS (i.e., LNF, PSF, NWF, ORF) significantly discriminated among students identified a priori at risk of EBD categorized as low average, average, and above average on the basis of WRMT-R Total scores. Ranking of the first grade DIBELS discriminant function coefficients revealed that the best discriminators across the groups were ORF and LNF. As to kindergarten DIBELS (i.e., ISF, LNF, PSF, and NWF), LNF is the best discriminator among students identified as at risk of EBD.

In first graders, the DIBELS were statistically significant discriminators of group membership for students identified as at risk of EBD on WRMT-R Total scores. ORF and LNF had the largest standardized

discriminant function coefficients. They made the greatest unique contribution to predicting group membership. For the first grade students, all DIBELS measures had moderate to high associations with the WRMT-R Total Reading cluster. This finding was consistent with Kaminski and Good (1996) although these authors found fewer positive associations between the DIBELS and first grade criterion measures than in kindergarten. Findings were also consistent with Tobin (2000), who found that the subtests of the *Woodcock Diagnostic Reading Battery* and *Test of Oral Reading Fluency* correlated positively with DIBELS PSF (.44 to .70) and NWF (.55 to .88) for regular education first grade students.

Our kindergarten results were partially consistent with previous research (Elliot et al., 2001; Hintze et al., 2002; Kaminski & Good, 1996; Speece, Mills, Ritchey, & Hillman, 2003) which showed statistically significant, positive validity correlations between the kindergarten DIBELS and various other criterion measures (e.g., *Comprehensive Test of Phonological Awareness*, *Metropolitan Readiness Test*, *Level*, *Woodcock-Johnson Psycho-Educational Battery-Revised*). There are four likely explanations for the kindergarten results. First, most previous studies have used a “level” rather than a “point” estimate for the DIBELS. Level estimates refer to the average of DIBELS scores over repeated administrations whereas “point” refers to one administration at a given time. Perhaps the kindergarten student’s scores at one point in time under a unique set of conditions resulted in an attenuated relationship between the DIBELS and WRMT-R Total score. Second, the homogeneous nature of the students in the present study may have restricted the range of scores. Previous studies have relied on heterogeneous general education students for their samples. In this study, students were selected who met stringent criteria for identification of at-risk EBD status. The homogeneity of the groups may have functioned to restrict the variance in such a way that it attenuated the correlations between DIBELS measures and the WRMT-R Total scores. Third, the timing of the test, DIBELS, was administered early in the fall. In the early months of kindergarten, students may not have yet fully developed the prerequisite reading skills to a degree to be discriminated by ISF, PSF, LNF and NWF. For example, Hintze et al., (2001) indicated that early in kindergarten students are just beginning to develop and fine tune the phonological awareness skills assessed by PSF. Speece et al., (2003) found that the NWF may not be a valid measure in kindergarten until the spring. A fourth consideration is the lack of predictive power associated with the low to moderate reliability estimates for the kindergarten DIBELS. Specifically, ISF possesses an alternate form reliability estimate of .72, PSF ranges from .60 to .88, LNF ranges from .80 to .93, and PSF ranges from .84 to .88, all considered low for educational decision making.

## Implications for Practice

The concurrence of reading problems with emotional and behavior problems has long been evidenced by teachers and is documented in the research literature. Interestingly a concurrent development of a three-tier model for assessment and intervention is found as the prevailing framework separately in both the reading and behavior disorders literature: the three-tier model as schoolwide best practice for level one; screening and subsequent identification of students at risk in level two; and individualized assessment and intervention in

level three. However, assessments and interventions rarely address reading and behavior together, instead discussing these as mutually exclusive categories of classroom problems. This study indicates that there is utility in examining the intersection of identification of students at risk for emotional and behavioral problems and students at risk for reading problems. The DIBELS are efficient and effective for early screening and identification of at-risk students before they become well entrenched in reading failure and on a path to negative emotional and behavioral outcomes. Early screening and identification is often proposed as a means of prevention of both reading and behavior problems; but among subgroups of at-risk students, those at risk for EBD, the identification and screening of reading problems is seldom discussed. DIBELS can reliably discriminate reading characteristics specific to this vulnerable subgroup of students as early as fall of the first grade year.

As with any assessment, DIBELS should not be used as the sole criterion to make diagnostic or intervention decisions (Kaminski & Good, 1996). To do so may render false positives, leading educational personnel to inaccurately identify early elementary students as being “at risk” for reading difficulties (Hintze et al., 2002) when in fact they are not. However, providing additional reading instruction, especially for those also identified as “at risk” for emotional and behavioral problems, is unlikely to do harm. The usefulness of the DIBELS for discriminating ability characteristics of students at risk of EBD is good news for schools. Service providers (e.g., teachers) in local settings can administer this type of assessment quickly and easily receiving reliable prediction of risk for reading failure thereby potentially reducing the risk of subsequent or co-occurring behavioral and emotional difficulties. Early identification can then lead instructional personnel to develop a stronger orientation for early intervention.

### Limitations and Future Research

The results of this study must be considered in the context of limitations in sampling and measurement. First, our sample size was small. As is often the case, replication would strengthen confidence in the findings. A second limitation relates to group formation. Rather than form two groups consisting of poor and good readers, three groups were a better fit to the data. Most studies rely on differentiating between poor and good performance groups. Replicating these findings will rely on validating the three group membership. Future research should consider larger samples of students at risk for or identified as EBD. In addition, longitudinal studies might better follow the same sample from kindergarten through first grade and beyond. This would permit a more useful barometer of the predictive utility of the DIBELS in later reading ability, rather than using two mutually exclusive groups. Finally, because the DIBELS data relied only on one administration, it was not possible to calculate any reliability estimates (e.g., test-retest) for our sample as is customarily done. Future studies will attempt to use alternate forms to calculate alternate-form reliability.

### Conclusion

An increasing number of children arrive at school with a plethora of risk factors and needs that are likely to affect their academic survival (Morrison, Walker, Wakefield, & Soldberg, 1994). This trend will likely

continue given that today’s children are more at risk of social, emotional, behavioral, and academic problems than ever before (Knoff, Curtis, & Batsche, 1997). The increase in the at-risk status of children has occurred in the context of an educational climate that demands efficient and data-driven decision making that is aligned to prevention and intervention while being linked to assessment (Hintze et al., 2002). The ability to identify reading groups as early as the fall of first grade provides tremendous possibility for early intervention.

Research tells us that when young children lag behind their peers in reading, they are unlikely to catch up without strategic, targeted, and systematic instruction in key skills required for reading success. Further, there is comorbidity between reading problems and emotional and behavioral problems. Accurate, early identification of reading problems in students who are at risk of EBD may be the mitigating link between reading failure and the development of emotional and behavioral difficulties. Our results indicate that early identification is certainly possible and that this reading-behavior predictive data can assist in the targeting of individuals in need for additional services to prevent possible future school failure and drop out. A tool, such as the DIBELS, for discriminating reading characteristics of students at risk of EBD has a high utility for schools as they strive to better understand and intervene in the intersection between reading failure and problem behavior through evidence-based assessment practices.

---

## References

- Brier, N. (1995). Predicting antisocial behavior in youngsters displaying poor academic achievement: A review of risk factors. *Developmental and Behavioral Pediatrics, 16*, 271-276.
- Buck, J., & Torgesen, J. K. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test*. Tallahassee, FL: Florida Center for Reading Research.
- Coutinho, M. J. (1986). Reading achievement of students identified as behaviorally disordered at the secondary level. *Behavioral Disorders, 11*, 200-207.
- Coleman, M., & Vaughn, S. (2000). Reading interventions for student with emotional and behavioral disorders. *Behavioral Disorders, 25*, 93-104.
- Cullinan, D., & Epstein, M. H. (2001). Comorbidity among students with emotional disturbance. *Behavioral Disorders, 26*, 200-213.
- Cullinan, D., Evans, C., Epstein, M. H., & Ryser, G. (2003). Characteristics of emotional disturbance of elementary school students. *Behavioral Disorders, 28*, 94-110.
- Duarte-Silva, A. P., & Stam, A. (2004). Discriminant analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistic* (pp.277-317). Washington, DC: American Psychological Association.
- Elliot, J., Lee, S., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills—modified. *School Psychology Review, 30*, 33-49.
- Good, R. H., Gruba, G. G., & Kaminski, R. A. (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS). In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 679-700). Washington, DC: National Association of School Psychologists.

- Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27*, 45-56.
- Gottfredson, D. C. (1981). School and delinquency. In S. E. Martin, L. B. Sechrest & R. Redner (Eds.), *New directions in the rehabilitation of criminal offenders*. WA: National Academy.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hawkins, J. E., & Lishner, D. M. (1987). Schooling and delinquency. In E. H. Johnson (Ed.), *Handbook of crime and delinquency prevention* (pp. 245-278). New York: Greenwood.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2002). *Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Awareness*. Amherst, MA: University of Massachusetts.
- Kaminski, R., & Good, R. I. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.
- Knoff, H. M., Curtis, M. J., & Batsche, G. M. (1997). The future of school psychology: Perspectives on effective training. *School Psychology Review, 26*, 93-103.
- Kutach, K., Duchnowski, A. J., & Friedman, R. M. (2005). The system of care 20 years later. In M. H. Epstein, K. Kutasch, & A. J. Duchnowski (Eds.), *Outcomes for children and youth with emotional and behavioral disorders and their families: Programs and evaluation of best practices* (2<sup>nd</sup> ed.) (pp. 3-22). Austin, TX: PRO-ED.
- Lane, K. L. (2004). Academic instruction and tutoring interventions for students with emotional and behavioral disorders: 1990 to the present. In R. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of Research in Emotional and Behavioral Disorders*. (pp. 462-486). New York: Guilford Press.
- Mooney, P., Epstein, Reid, R., & Nelson, R. (2003). Status of and trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education, 24*, 473-287.
- McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic failure, and school climate: A critical review. *Journal of Emotional and Behavioral Disorders, 8*, 130-141.
- Morrison, G. M., Walker, D., Wakefield, P., & Soldber, S. (1994). Teacher preferences for collaborative relationships: Relationship to efficacy for teaching in prevention-related domains. *Psychology in the Schools, 31*, 221-231.
- Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities. *Remedial and Special Education, 27*, 276-292.
- Roy, J., & Bargman, R. E., (1958). Tests of multiple independence and the associated confidence bounds. *Annals of Mathematical Statistics, 29*, 491-503.
- Ruhl, K. L., & Berlinghoff, D. H. (1992). Research on improving behaviorally disordered students' academic performance: A review of the literature. *Behavioral Disorders, 17*, 178-190.
- Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)*. Eugene, OR: University of Oregon.
- Silberberg, N. E., & Silberberg, M. C. (1971). School achievement and delinquency. *Review of Educational Research, 41*, 17-33.
- Speece, D., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education, 36*, 223-233.
- Tobin, K. J. (2000, July 7). *The effect of direct instruction and prior phonological awareness training on the development of reading skills in first grade*. Retrieved October 29, 2007 from the World Wide Web: <http://www.sraonline.com/download/DI/Research/Reading/HorizonsResearch.pdf>
- Trout, A. L., Nordness, P. D., Pierce, C. D., & Epstein, M. H. (2003). Research on the academic status of children and youth with emotional and behavioral disorders: A review of the literature from 1961-2000. *Journal of Emotional and Behavioral Disorders, 11*, 198-210.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO-ED.
- Wagner, M., Kutash, M., Duchnowski, A. J., Epstein, M. H., & Sumi, W. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13*(2), 79-96.
- Walker, H., & Severson, H. (1990). *Systematic screening for behavior disorders*. Longmont, CA: Sopris West.
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests-Revised*. Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., Johnson, M. B., Mather, N., McGrew, K. S., & Werder, J. K. (1991). *Woodcock-Johnson Psycho-educational Battery—Revised*. Itasca, IL: Riverside.

---

## Authors

**Jorge E. Gonzalez**, Ph.D., is an Assistant Professor in the School of Psychology at Texas A&M University. His research interests are literacy and language development in young children.

**Kimberly J. Vannest**, Ph.D., is an Assistant Professor in the School of Psychology at Texas A&M University. Her research interest is interventions for students with emotional and behavioral disorders.

**Robert Reid**, Ph.D., is a Professor in the College of Education and Human Services at the University of Nebraska-Lincoln. His research interests are self-regulation and strategy instruction for students with learning disabilities and attention deficit hyperactivity disorder.

---