School Psychology Review,
2001, Volume 30, No.1

GENERAL ARTICLES

A Reliability and Validity Study of the Dynamic Indicators of Basic Early Literacy Skills–Modified

Jacquelyn Elliott, Steven W. Lee, and Nona Tollefson
University of Kansas

Abstract. This study examined the psychometric properties of a set of preliteracy measures modified from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) with a sample of 75 kindergarten students. The modified battery (called DIBELS-M) includes measures of Letter Naming Fluency, Sound Naming Fluency, Initial Phoneme Ability, and Phonemic Segmentation Ability. These measures were assessed through repeated administrations in 2-week intervals at the end of the kindergarten year. Interrater reliability estimates and coefficients of stability and equivalence for three of the measures ranged from .80 to the mid .90s with about one-half of the coefficients above .90. Correlations between DIBELS-M scores and criterion measures of phonological awareness, standardized achievement measures, and teacher ratings of achievement yielded concurrent validity coefficients ranging from .60 to .70. Hierarchical regression analysis showed that the four DIBELS-M measures accounted for 73% of the variance in scores on the Skills Cluster of the Woodcock-Johnson Psychoeducational Battery—Revised (WJ-R). The results of the analysis support the use of the DIBELS-M measures for identification of kindergarten students who are at-risk for reading failure and for progress monitoring. The contributions of the study, including psychometric analysis of the DIBELS-M with a new sample and formation of composite scores, are discussed in relation to the extant literature.

Within the current climate of school reform initiatives, alternative assessment methods have been widely promoted in the field of education (Bagnato, Neisworth, & Munson, 1989; Miller, 1995). This pressure for changes in the nature of assessments has grown from the assumption that good assessment is an integral part of good instruction and that conventional child assessments do not yield instructionally relevant information (Herman, Aschbacher, & Winters, 1992). The pressure for changes in assessment practices for very young children has been more pronounced than for older school-aged students because of the limited technical adequacy of current developmental and readiness instruments (Drieling & Copeland, 1988; Meisels, Wiske, & Tivnan; 1984; Tramontana, Hooper, & Selzer, 1988).

This emphasis on the inseparability of curriculum and assessment and the premise that assessment activities should contribute to instructional improvement has raised major challenges for educational measurement. Many of the alternative assessment methods have not been evaluated for technical adequacy or have yielded low reliability and/or validity estimates

Address all correspondence concerning this article to Steven Lee, Ph.D., University of Kansas, Psychology and Research in Education, 621 Pearson Hall, 1122 West Campus Rd., Lawrence, KS 66045. E-mail: swlee@ku.edu.

(Koretz, Stecher, Klein, & McCaffrey, 1994; Linn & Baker, 1996; Moss, 1992; Shavelson, Baxter, & Pine, 1992). Performance-based assessment, student portfolios, authentic assessment, curriculum-based assessment, and curriculum-based measurement have all been described in the literature and used by school psychologists as forms of alternative assessment (Shinn, 1995). However, considerable variability remains in the technical attributes and feasibility of implementation for these instruments (Burstein, 1994; Lane, Stone, Ankenmann, & Liu, 1992).

Curriculum-based measurement (CBM) is one of the few alternative forms of assessment where an impressive body of data supports the technical adequacy and the practical application of these techniques in the assessment of young children (Deno, 1985; Deno & Fuchs, 1987; Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 1997; Germann & Tindal, 1985; Marston & Magnusson, 1985; Shinn, 1989). School psychologists have used curriculum-based measurement as a form of performance assessment that features measurement of student proficiency across core areas of the curriculum. Curriculum-based measures were developed and standardized in order to facilitate frequent, ongoing assessment of basic skills and formative evaluation of student progress. The technical adequacy of CBM measures for late first through sixth grades has been well documented in the literature (Fuchs, 1994; Fuchs & Deno, 1994; Marston & Magnusson, 1985; Shinn, Good, Knutson, & Tilly, 1992; Stoner, 1992; Tindal, 1993).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

The original Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good, Kaminski, Laimon, & Johnson, 1992; Good, 1994) are a set of 10 brief measures designed for progress monitoring and early identification of children with reading problems. The DIBELS measures were originally conceptualized as downward extensions of CBM reading probes. Like many other curriculum-based measures, both point and level estimates of performance have been used. The point estimate score describes the student's performance on a single measure, whereas the level estimate score is based on the average of all repeated measures for a given task during a specified data collection period.

The DIBELS evaluate a set of early literacy skills identified in the literature as directly related to and facilitative of later reading competence. Student knowledge of letter names, sound-symbol relationships, and phonemic awareness in kindergarten have all been identified as important predictors of later literacy (Blachman, 1984, 1989; Felton & Wood, 1989; Stahl & Murray, 1994; Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976; Torgesen, Morgan, & Davis, 1992). The DIBELS incorporates brief measures of each of these important abilities along with other potentially important general language and associated abilities. DIBELS measures of language development include: Story Retell, Picture Description, and Picture Naming Fluency. DIBELS tasks designed to measure knowledge of alphabetic print are: Letter Naming and Sound Naming Fluency. DIBELS measures of phonological awareness include: Rhyming Fluency, Blending Fluency, Onset Recognition, Initial Sound Fluency, and Phonemic Segmentation Fluency.

In a recent study published by the University of Oregon research group (Kaminski & Good, 1996), DIBELS Letter Naming Fluency, Phonemic Segmentation Fluency, and Picture Naming Fluency measures were evaluated for a kindergarten cohort group of 18 children. The one-week alternate form reliability coefficients for point estimate scores were .93 for Letter Naming and .88 for Phonemic Segmentation. The alternate form reliability coefficient based on level estimates was .99 for Letter Naming and for Phonemic Segmentation. Criterion-related validity coefficients ranged between .58 and .90 for Letter Naming level estimate scores and between .63 and .73 for Phonemic Segmentation level estimate scores. Point estimate scores for both measures yielded similar but slightly weaker correlation coefficients. Coefficients for Picture Naming Fluency were weaker than for the other two

DIBELS tasks (e.g., point estimate alternative forms reliability coefficient for Picture Naming Fluency was .77). Based on the psychometric analysis of the data, the authors concluded these DIBELS measures provide a reliable and valid indicator of children's progress toward the acquisition of early literacy skills.

The Present Study

This study investigated the technical adequacy of four selected DIBELS measures in identifying kindergartners who are at-risk for reading failure. These measures are Letter Naming Fluency, Sound Naming Fluency, Initial Phoneme Ability, and Phonemic Segmentation. Kaminski and Good (1996) conducted their investigation with 18 participants. The present study provides an extension of their reliability and validity studies of the DIBELS by drawing a larger and more diverse sample from an urban school district in a moderate-sized midwestern city. Letter Naming Fluency (LNF) and Sound Naming Fluency (SNF) were included in the present study because letter-sound correspondences have been widely trained and measured with kindergarten children (Ball & Blachman, 1991; Bradley & Bryant, 1985; Vellutino & Scanlon, 1987). In addition, many models of reading acquisition specify the importance of some letter-sound correspondence knowledge in providing explicit cues to young readers as they decode unfamiliar words (Ehri & Wilce, 1985; Marston & Magnusson, 1988). The remaining two phonemic awareness measures (Initial Phoneme Ability and Phonemic Segmentation Ability) were renamed Initial Phoneme Ability (IPA) and Phonemic Segmentation Ability (PSA) in order to distinguish them from the original DIBELS measures (called Initial Sound Fluency and Phonemic Segmentation Fluency). These two measures were included in the present study to emphasize accuracy over fluency as well as to correct for floor effects found by Good et al. (1992). On both of these measures, simpler stimulus words were selected from the kindergarten reading text used in the participating school district, and children were given additional time for responding. Based on these modifications, the measures were viewed more as measures of ability in the sense that the child's phonemic knowledge was tapped without their performance being timed as on the fluency tasks. The four measures included in the present study are described in detail in the instrument section of the paper and referred to as the DIBELS-M.

Method

Participants

Kindergarten children ($n = 75$) from four classrooms in three elementary schools in a moderate-sized midwestern city participated in the study. The sample was representative of both the district and national school-age populations. Forty-one students were male and 34 students were female. Approximately 37% of the sample were nonwhite. Approximately one third of the sample ($n = 27$) were eligible for free or reduced lunches. Parent permission was given for all kindergarten children in the four classrooms to participate in the five assessment sessions. All children who participated had passed hearing and vision screens. A small percentage of children enrolled in these kindergarten classes had already been selected for special instructional programming (e.g., speech and language, remedial reading, learning disabilities services) and were included in the assessments along with their classmates.

Instruments

Students completed five assessments: the DIBELS-M battery; the Woodcock-Johnson Psychoeducational Achievement Battery—Revised (WJ-R; Woodcock & Johnson, 1989, 1990), Broad Reading and Skills clusters; The Test of Phonological Awareness—Kindergarten form (TOPA; Torgesen & Bryant, 1994); and the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990). The Developing Skills Checklist (DSC; CTB Macmillan/McGraw-Hill, 1990) was administered early in the school year by staff members in the participating school district. The DIBEL-M measures were the predictor measures and the WJ-R, TOPA, DSC, and an informal teacher rating questionnaire were the criterion measures. The K-BIT scores were used to control for differences in ability in the regression analyses.

Predictor measures. The following four DIBELS-M measures were administered:

1. Letter Naming Fluency (LNF). Children were shown a card containing upper and lower case letters printed along 11 rows, with 10 letters per row. Children were asked to name as many letters as they could in 60 seconds, proceeding from left to right and down the card. The three alternate forms of the LNF test were part of a set developed by the research group of Good et al. (1992).

The examiner had an identical stimulus card and recorded whether letters were identified correctly or incorrectly. If the child delayed longer than 3 seconds on a particular letter, the examiner recorded an error and supplied the child with the name of the letter. The LNF score was the number of letters correctly identified in the 60-second trial.

2. Sound Naming Fluency (SNF). The three forms of this test were also taken from the set of cards developed by Good et al. (1992) for the LNF test. For the SNF tasks, children were instructed to give the sound made by each letter. Children proceeded through as many letters as possible in 60 seconds. If the child did not offer a response within 3 seconds, the examiner supplied the correct letter sound and recorded the item as an error. The SNF score was the number of letter sounds correctly identified in the 60-second trial.

3. Initial Phoneme Ability (IPA). This test was structured in a manner similar to the Initial Sound Fluency task used by Good et al. (1992). Some items from the Initial Sound Fluency task were incorporated in the IPA test along with high frequency CVC (consonant-vowel-consonant) words taken from the 1993 Macmillan/McGraw-Hill reading series used in the participating school district.

Each form of the IPA test consisted of 10 CVC words. The examiner read the words to the children and instructed them to repeat the initial sound heard in the word following a two-item trial. The examiner recorded whether children identified the initial phoneme correctly or incorrectly. If a child delayed a response longer than 5 seconds, the examiner recorded an error in the data record. The IPA score was based on the total number of initial phonemes correctly identified in accordance with scoring procedures for Good et al.'s (1992) Initial Sound Fluency measure.

4. Phonemic Segmentation Ability (PSA). This test was very similar to Good et al.'s (1992) Phonemic Segmentation Fluency test, and it differed from the modified IPA task described above only in that children were asked to supply *all* of the phonemes composing the stimulus word. Each form of the PSA test consisted of 10 CVC words for a total of 30 phonemes. Again, some items from the DIBELS Phonemic Segmentation Fluency test were retained and incorporated in the revised PSA tests along with high frequency CVC words from the Macmillan series.

The examiner read the word and instructed the child to say the separate sounds that composed the word. The examiner recorded correct phonemes and incorrect pho-

Table 1
Intercorrelations Between the DIBELS-M Subtests

Measure	LNF	SNF	IPA	PSA
Letter Naming Fluency (LNF)	1.0	.71	.42	.38
Sound Naming Fluency (SNF)		1.0	.56	.50
Initial Phoneme Ability (IPA)			1.0	.67
Phonemic Segmentation Ability (PSA)				1.0

Note. All correlations are significantly greater than 0.

Table 2
Factor Loadings for the Principal Components Factor Analysis

Measure	Factor Loads	
	Factor 1	Factor 2
Letter Naming Fluency (LNF)	.223	.918
Sound Naming Fluency (SNF)	.394	.834
Initial Phoneme Ability (IPA)	.846	.394
Phonemic Segmentation Ability (PSA)	.923	.226

Note. All correlations are significantly greater than 0.

nemes produced for each word. If children took longer than 15 seconds to complete their responses on a particular word, the examiner recorded errors for the remaining phonemes. The PSA score was based on the total number of phonemes correctly identified.

Individual DIBELS-M scores were combined based on patterns of the intercorrelations among the DIBELS-M measures into three composite scores designated "Fluency," "Ability," and "Total." Intercorrelations among scores on the four DIBELS measures are reported in Table 1. A Principal Components factor analysis yielded a two factor solution. Table 2 reports the factor loadings for a two factor structure. Factor 1, labeled Ability, was defined by the IPA and PSA measures; Factor 2, labeled Fluency, was defined by the LNF and SNF measures. Following varimax rotation, Factor 1 accounted for 43% of the variance and Factor 2 accounted for 42% of the variance.

To compute these composite scores, scores for each of the four measures were first converted to standard scores with a mean of 100 and a standard deviation of 15. The Fluency score was derived by averaging student's standard scores for LNF and SNF. The Ability score was derived by averaging student's standard scores for IPA and PSA. The Fluency measures placed more emphasis on the speed of correct responses whereas the two Ability measures were designed to reflect accuracy of responses over the speed of responses. Finally, the Total score was computed by averaging the subject's Fluency and Ability scores.

Studies of the technical characteristics of the DIBELS (Good et al., 1992) indicated acceptable levels of reliability and predictive validity for the scales. Predictive validity coefficients between kindergarten DIBELS measures and the first grade criterion measures were all positive and significant. For Letter Naming Fluency, predictive validity coefficients ranged from .68 with the Stanford Diagnostic Reading Test to .92 with the Teacher Rating Scale. Correlations between Phonemic Segmentation Fluency and the criterion measures ranged from .89 with the Teacher Rating Scale to .60 with the Stanford Diagnostic Reading Test. Validity coefficients for Picture Naming Fluency ranged from .81 with teacher ratings to .78 with the CBM-Reading level estimate. For all of these measures, higher means were found in first grade than in kindergarten.

In a second study, Good et al. (1992) administered Rhyming Fluency, Initial Sound Fluency, Blending Fluency, and Phonemic Segmentation measures to 100 kindergarten-aged children four times over the latter half of the school year. Test-retest reliability coefficients ranged from the low .70s for Phonemic Segmentation, Initial Sound, and Rhyming Fluency to .53 for Blending Fluency.

Criterion measures. In addition to the four DIBELS-M measures, three achievement measures, a teacher rating scale, and brief intelligence test were used as criterion measures:

1. The Woodcock-Johnson Psycho-Educational Achievement Battery—Revised (WJ-R) Broad Reading and Skills clusters

were selected as criterion measures because of their strong technical qualities and their widespread use in the identification of reading disabilities among school-aged children. In a survey of the current use of assessment instruments among practicing school psychologists, the WJ-R was identified as the most frequently used achievement measure (Wilson & Reschly, 1996). The WJ-R Broad Reading cluster is based on a letter-word recognition subtest and on a simplified reading comprehension subtest. The Skills cluster consists of a letter-word identification subtest, a drawing and spelling subtest, and an applied problems subtest.

Internal consistency reliability coefficients for the WJ-R are in the mid-90s for the achievement cluster scores and exceed .90 for the separate subtest scores (McGrew, Werder, & Woodcock, 1991). Shull-Senn, Weatherly, Morgan, and Bradley-Johnson (1995) reported coefficients of stability for the WJ-R Broad Reading cluster exceeded .90 at Grades 1, 3, and 5. Coefficients for the letter-word identification and passage comprehension subtests exceeded .90 at Grade 1. Content, concurrent, and construct validity studies, which have included preschool samples, have demonstrated that the WJ-R assesses multiple facets of important academic skills (Cummings, 1995). The WJ-R has also been used as the criterion measure in evaluating the concurrent validity of a number of recently developed instruments, for example, the Wechsler Individual Achievement Test (Martelle & Smith, 1994) and the Detroit Test of Learning Aptitude-Primary (Dunn, 1992). The WJ-R has also been routinely used as a dependent measure of reading achievement for young children (Burns, Collins, & Pausell, 1991; Fletcher et al., 1994; Jackson, Donaldson, & Mills, 1993; Stipek, Feiler, Daniels, & Milburn, 1995).

2. The Test of Phonological Awareness (TOPA) was designed to identify children who could benefit from phonological awareness training in order to help them prepare for reading instruction in first grade. The TOPA uses a group-administered format to measure children's ability to isolate initial phonemes in spoken words. Torgesen and Bryant (1994) reported a correlation of .62 between kindergarten TOPA scores and end of first grade performance on the Word Analysis subtest from the Woodcock Reading Mastery Test. In addition, children who scored in the lowest quartile on the TOPA in kindergarten also scored below the median on the first grade word analysis subtest.

3. Teacher Rating Questionnaire (TRQ). Teachers rated students' present level of pre-reading at the end of the school year. The response scale ranged from "well below average" to "well above average" with scores scaled between 1 and 5. The teacher rating questionnaire was adapted from one developed by Share, Jorm, MacLean, and Matthews (1984). No reliability or validity data are reported by the authors.

4. The Developing Skills Checklist (DSC) was designed as an individually administered criterion-referenced screening test to help teachers plan instructional programs for preschool- and kindergarten-aged children. Abilities evaluated by the DSC include: mathematics concepts, logical concepts, memory, gross motor, fine motor, auditory, print concepts, and writing and drawing concepts. The DSC Pre-Reading Total Score was used for statistical analysis in the present study. Internal consistency reliability coefficients ranged between .81 and .95 for all DSC scales. No test-retest reliability or predictive validity estimates were reported by the authors. Although evidence for construct validity was weak, the data suggest adequate attention to the content validity of the DSC (Clark, 1995).

5. The Kaufman Brief Intelligence Test (K-BIT) is composed of two subtests measuring verbal (i.e., Expressive Vocabulary) and nonverbal (i.e., Matrices) abilities. Split-half and test-retest reliabilities are reported by age level and range from .86 to .98 for the Vocabulary subtest, from .74 to .95 for the Matrices subtest, and from .88 to .98 for the composite score. Moderate to high correlations with other intelligence tests have been reported. Concurrent validity studies between the K-BIT and more comprehensive individual tests of intelligence, including the Wechsler Intelligence Scale for Children—Third Edition, a short form of the Stanford-Binet—Fourth Edition, and the Wechsler Intelligence Scale for Chil-

dren—Revised, have consistently supported the use of the K-BIT as a screening instrument for general cognitive abilities (Prewett, 1992; Prewett & McCaffery, 1993; Smith, 1992). The K-BIT was administered in the present study to examine the relationship between the DIBELS-M measures and general cognitive ability. K-BIT composite scores were also used as a statistical control for general cognitive ability in hierarchical multiple regression analyses examining the relationship between the DIBELS-M and the earlier mentioned criterion measures.

Procedures

Students were tested individually every 2 weeks for a period of 9 weeks (i.e., four assessments) at the end of their kindergarten year. In addition, a group administration of the TOPA was given between the first and second DIBELS-M assessments to groups of five or six children. The assessments were conducted by six practicing school psychologists including the first author, following a brief training session on administration and scoring of the DIBELS-M. The data collection process is described in Table 3.

An ancillary objective of this study was to demonstrate the equivalence of these alternate DIBELS-M forms. Each kindergartner was administered one of three parallel forms of the four DIBELS-M measures. The administration of Forms A and B were alternated during the first assessment session. During the third and fourth assessment sessions, Form C of each DIBELS-M measure was administered to all students. This format allowed analysis of two types of reliability indices, equivalent forms (with time intervening) and test-retest. The K-BIT and the WJ-R were administered in a counterbalanced fashion during the fourth individual assessment session. Kindergarten teachers completed the Teacher Rating Questionnaire during the final week of the study.

Data Analysis

To examine interrater reliability, the investigator rescored data from 50 participants for the four DIBELS-M measures using audiotape recordings of the sessions. The audiotapes used for evaluating interrater reliability were selected randomly from those sessions in which (a) the investigator had not served as the examiner and (b) the audiotape was sufficiently clear to permit rescoring of the child's performance. Percent agreement reliability estimates were computed on the item responses. These coefficients are reported in the Results section.

Correlations were computed between the level estimates (average scores) derived from each of the DIBELS-M measures and each achievement-related criterion measure. Simultaneous and hierarchical multiple regression analyses were performed separately for each criterion measure and were used to understand which of the DIBELS-M measures made unique contributions to the prediction of the criterion reading measures. Three regression designs were used: Design 1 was a simultaneous multiple regression analysis, Design 2 and Design 3 were hierarchical multiple regres-

Table 3
Schedule of Data Collection

Session 1	Session 2	Session 3	Session 4	Session 5
Week 1	Week 2	Week 3	Week 5	Week 7
DIBELS-M Forms A & B	TOPA Group Testing	DIBELS-M Form C	DIBELS-M Form C	K-BIT WJ-R

DIBELS-M = Dynamic Indicators of Basic Early Literacy Skills-Modified
K-BIT = Kaufman Brief Intelligence Test
TOPA = Test of Phonemic Awareness
WJ-R = Woodcock-Johnson Psycho-Educational Achievement Battery—Revised

Table 4
Means and Standard Deviations for the DIBELS-M Point and Level Scores, Criterion Achievement Measures and Ability Measures (n = 75)

Predictor Measures	"Point"			"Level"
Measure	Session 1	Session 2	Session 3	Combined
Letter Naming Fluency				
M	24.31	27.13	32.41	27.95
S.D.	15.99	15.54	17.38	15.29
Sound Naming Fluency				
M	7.12	10.55	11.99	9.88
S.D.	6.41	8.20	9.23	7.47
Initial Phoneme Ability				
M	5.43	6.01	6.85	6.10
S.D.	3.57	3.73	3.44	3.21
Phonemic Segmentation Ability				
M	11.33	13.91	15.87	13.70
S.D.	8.37	9.66	9.04	8.46

Criterion Measures	*M*	*S.D.*
Woodcock-Johnson Psychoeducational Achievement Battery*		
Broad Reading Cluster	91.73	14.79
Skills Cluster	95.11	13.00
Letter-Word Id. Subtest	91.37	15.27
Test of Phonological Awareness	94.60	13.47
Teacher Rating Questionnaire	3.49	0.91
Developing Skills Checklist	51.49	20.23

Ability Measure	*M*	*S.D.*
Kaufman Brief Intelligence Test		
Verbal Score	96.69	11.88
Nonverbal Score	97.65	9.27
Composite Score	96.84	9.74

*Standardized scores for the WJ-R are based on grade norms rather than age norms.

sion analyses. In all analyses, the criterion variables were the WJ-R Broad Reading Cluster score, the WJ-R Skills Cluster score, and the Teacher Rating Scale score. In the first hierarchical design, K-BIT scores were entered initially so that the unique contribution of the four DIBELS-M measures could be determined after accounting for differences in student abilities. In the second hierarchical analysis, TOPA scores were entered after K-BIT scores to determine the unique contribution of the DIBELS-M measures to predicting WJ-R scores and teacher ratings of achievement. In all of the regression analyses, DIBELS-M measures were entered into the regression equation in a block.

Results

Descriptive Statistics For the Predictive and Criterion Measures

Means and standard deviations for the point and level estimates for each of the three administrations of the DIBELS-M (the predic-

tor measures), the four criterion measures of achievement, and the K-BIT are presented in Table 4. Simple within-subjects analyses of variance confirmed significant increases in point scores over the three administrations of the DIBELS-M scales: Letter Naming Fluency (LNF): F (2, 148, n = 75) = 25.86, $p < .001$; Sound Naming Fluency (SNF): F (2, 148, n = 75) = 35.58, $p < .001$; Initial Phoneme Ability (IPA): F(2, 148, n = 75) = 10.30, $p < .001$; Phonemic Segmentation Ability (PSA): F (2, 148, n = 75)= 25.41, $p < .001$.

Reliability Estimates for the DIBELS-M

Three types of reliability estimates were computed: interrater reliability, test-retest reliability, and alternate forms reliability. The results of these analyses are summarized in Table 5. The item-by-item interrater percent agreement fell generally in the mid to upper .80s for individual and composite DIBELS measures. Item-by-item ratings were strongest (.94) for LNF and weakest (.82) for SNF. Correlations between total scores assigned by different raters were in the mid to high .90s.

Test-retest reliability was evaluated by correlating DIBELS-M scores from the second and third testing sessions. Two weeks intervened between the testing sessions. Reliability estimates are based upon repeated administrations of Form C of the DIBELS-M. Coefficients of stability were .90 for LNF, .83 for SNF, .74 for IPA, and .85 for PSA. Higher correlations were observed for two of the three DIBELS-M composite scores (i.e., .93 for Fluency and Total and .86 for Ability).

Alternate forms reliability was estimated by correlating the scores from the first and second testing sessions. During the first testing session, half the students were administered Form A of the DIBELS-M and half were administered Form B. During the second testing session, all students were administered Form C. As shown in Table 5, these coefficients, with the exception of IPA, ranged from .80 to .91.

All reliability estimates, with the exception of the coefficients for IPA, were .80 or higher. Overall, estimates of interrater reliability were in the high .80s to .90s. As expected, coefficients of stability were higher than coef-

Table 5
Reliability Estimates for the DIBELS-M Measures

Measure	Interrater Reliability	Test Retest	Equivalent Forms
	N = 50	N = 75	N = 75
Letter Naming Fluency (LNF)	.94	.90	.80
Sound Naming Fluency (SNF)	.82	.83	.82
Initial Phoneme Ability (IPA)	.89	.74	.64
Phonemic Segmentation Ability (PSA)	.87	.85	.84
Composite Scores			
Fluency	.90	.93	.87
Ability	.88	.86	.81
Total	.89	.93	.91

Note. All correlations are significantly greater than 0. The Fluency Composite is the combined scores for Letter Naming Fluency and Sound Naming Fluency. The Ability Composite is the combined scores for the Initial Phoneme Ability and Phonemic Segmentation Ability.

ficients of stability/equivalence. However, the magnitude of the coefficients of stability/equivalence indicates that most of the variance in students' scores could be attributed to true differences in the abilities measured rather than to errors of measurement.

Concurrent Validity of the DIBELS-M

Correlations between "level" estimates (average scores over repeated administrations) and each of the criterion measures were used to estimate the concurrent validity of the DIBELS-M. These correlations are reported in Table 6. Level estimates for the DIBELS-M showed the strongest correlations with scores on Skills Cluster of the WJ-R and the Developing Skills Checklist. Correlations between the DIBELS-M measures and the criterion achievement measures explained, on average, between 35% and 40% of the variance in scores on these two achievement measures. Correlations with the WJ-R Skills cluster were .75 for LNF, .72 for SNF, .64 for IPA, and .60 for PSA. Correlations with scores on the Developing Skills Checklist showed the same pattern but were slightly lower in magnitude.

The fluency measures (LNF and SNF) were more strongly associated with the criterion measures than with the ability measure (IPA and PSA). The average correlation between the fluency measures and the achieve-

Table 6

Concurrent Validity of the DIBELS-M Level Estimate (Average) Scores

Measure	Individual				Composite		
	LNF	SNF	IPA	PSA	Fl	Ab	Tot
Achievement Measures							
Woodcock-Johnson Psychoeducational Achievement Battery							
Broad Reading	.63	.58	.42	.44	.64	.46	.62
Skills Cluster	.75	.72	.64	.60	.79	.66	.81
Letter-Word Id.	.71	.62	.47	.45	.71	.48	.67
Teacher Rating Questionnaire							
Prereading	.63	.62	.46	.53	.67	.52	.67
Developing Skills Checklist	.67	.69	.58	.54	.73	.59	.74
Test of Phonological Awareness	.50	.68	.62	.52	.63	.61	.69
Ability Measure							
Kaufman Brief Intelligence Test							
Verbal Score	.54	.55	.52	.36	.59	.47	.59
Nonverbal Score	.12	.25	.42	.40	.20	.44	.36
Composite Score	.42	.50	.57	.45	.49	.54	.58

Ab = Ability Composite Score.
Fl = Fluency Composite Score.
Tot = Total Composite Score.

ment measures was about .65. The average correlation between IPA and PSA and the achievement measures was in the low .50s. As expected, the Fluency Composite score showed the strongest correlations with the criterion achievement measures. However, the single strongest correlation was between the Total Score and the Skills Cluster on the WJ-R ($r = .81$). The concurrent validity indicated that the DIBELS-M measures explained no less than 16% of the variance and, in most cases, between 30% and 40% of the variance in students' scores on the achievement measures used as criteria.

Table 7
Model Summaries for the Simultaneous and Hierarchical Linear Regression
Criterion Variable: WJ-R Broad Reading Cluster

Model	R^2	*df*	*F*
Design 1			
DIBELS-M	.42	1,72	14.522**
Design 2			
K-BIT	.15	1,72	13.758**
DIBELS-M	.41	1,72	11.332**
Design 3			
K-BIT	.15	1,72	13.758**
TOPA	.23	1,72	11.816**
DIBELS-M	46	1,72	9.464**

Criterion Variable: WJ-R Skills Cluster

Model	R^2	*df*	*F*
Design 1			
DIBELS-M	.66	1,72	36 755**
Design 2			
K-BIT	.44	1,72	57 398**
DIBELS-M	.73	1,72	40.172**
Design 3			
K-BIT	.44	1,72	57 398**
TOPA	.53	1,72	41 871**
DIBELS-M	.75	1,72	33.431**

Criterion Variable: Teacher Rating Scale

Model	R^2	*df*	*F*
Design 1			
DIBELS-M	.49	1,72	18.181**
Design 2			
K-BIT	.24	1,72	24.088**
DIBELS-M	.52	1,72	16.513**
Design 3			
K-BIT	.24	1,72	24.088**
TOPA	.36	1,72	21.642**
DIBELS-M	.53	1,72	14 728**

**$p < .01$.

Table 8
Significant DIBELS Level and Composite Scores as Predictors of the Criterion Measures

Design 1 Level Scores Simultaneous Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Letter Naming Fluency	.435	3.34	.001
(WJ-R) Skills	Letter Naming Fluency	.438	4.37	.001
	Sound Naming Fluency	.219	1.99	.050
Teacher Rating	Letter Naming Fluency	.364	2.95	.004
	Phonemic Seg Ability	.363	2.79	.007
TOPA	Sound Naming Fluency	.504	3.76	.000

Design 2 Level Scores Hierarchical Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Letter Naming Fluency	.422	3.21	.002
(WJ-R) Skills	Letter Naming Fluency	.415	4.63	.000
	Phonemic Seg. Ability	.192	2.03	.046
Teacher Rating	Letter Naming Fluency	.341	2.85	.006
	Phonemic Seg. Ability	.365	2.90	.005
TOPA	Sound Naming Fluency	.492	3.59	.001

Design 3 Level Scores Hierarchical Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Letter Naming Fluency	.420	3.24	.002
(WJ-R) Skills	Letter Naming Fluency	.420	4.66	.000
	Phonemic Seg. Ability	.187	1.98	.050
Teacher Rating	Letter Naming Fluency	.353	2.99	.004
	Initial Phone Ability	-.362	-2.48	.015
	Phonemic Seg Ability	.352	2.83	.006

Design 4 Composite Scores Simultaneous Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Fluency Composite	.598	5.28	.001
(WJ-R) Skills	Fluency Composite	.615	7 125	.001
	Ability Composite	.280	3.24	.01
Teacher Rating	Fluency Composite	.566	5.18	.001
TOPA	Fluency Composite	.414	3.81	.01
	Ability Composite	.350	3.22	.01

Design 5 Composite Scores Hierarchical Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Fluency Composite	.569	4.82	.001
(WJ-R) Skills	Fluency Composite	.530	6.55	.001
Teacher Rating	Fluency Composite	.518	4.66	.001
TOPA	Fluency Composite	.382	3.38	.001
	Ability Composite	.309	2.64	.010

Table 8 (continues)

Table 8 (continued)

Design 6 Composite Scores Hierarchical Regression Analyses

Criterion Variable	Predictor	Beta	*T*	*p*
(WJ-R) Reading	Fluency Composite	.555	4.33	.001
(WJ-R) Skills	Fluency Composite	.522	5.95	.001
Teacher Rating	Fluency Composite	.462	3.87	.001

Note.
Design 1 Level Scores (simultaneous regression) CV = +[LNF+SNF+IPA+PSA].
Design 2 Level Scores (hierarchical regression): CV = [KBITS] +[LNF+SNF+IPA+PSA].
Design 3 Level Scores (hierarchical regression): CV = [KBITS] + [TOPA] + [LNF+SNF+IPA+PSA].
Design 4 Composite Scores (simultaneous regression) CV = +[Fluency Composite +Ability Composite].
Design 5 Composite Scores (hierarchical regression): CV = [KBITS] +[Fluency Composite +Ability Composite].
Design 6 Composite Scores (hierarchical regression): CV = [KBITS] + [TOPA] + [Fluency Composite +Ability Composite] where CV = criterion variable, LNF = Letter Naming Fluency, SNF = Sound Naming Fluency, IPA = Initial Phoneme Ability, PSA = Phonemic Segmentation Ability, Fluency Composite =LNF + SNF, Ability Composite = IPA + PSA, KBITS = composite score of the Kaufman Brief Intelligence Test, and TOPA = score for Test of Phonological Awareness.

Regression Analyses

To explore further the relationships between the four DIBELS-M measures and the four achievement measures, simultaneous and hierarchical regression analysis were conducted. The first hierarchical regression analysis is labeled "Design 1: simultaneous regression" in the subsequent tables. In this analysis, the composite score from the K-BIT was entered in the first block and the four DIBELS-M scores were entered simultaneously in the second block. The analysis permitted an examination of the unique contribution of the DIBELS-M measures to the prediction of the criterion achievement measure after accounting for the variance attributable to students' general mental ability. The second hierarchical analysis is labeled "Design 2: hierarchical regression" in subsequent tables. Here, K-BIT scores were initially entered into the regression equation, TOPA scores were the second variable entered, and the DIBELS-M measures were the third set of variables entered as a block. The adjusted R^2 values for the three regression analyses are reported in Table 7. The standardized beta weights for the DIBELS-M measures that were significant predictors of each of the three achievement measures (WJ-R Broad Reading and Skills Cluster, and Teachers' Ratings) are reported in Table 8.

Each of the four simultaneous regression analyses was significant. The set of four DIBELS-M measures explained significant proportions of the variance in the WJ-R Cluster and TOPA scores and teachers' ratings of student achievement. The set of DIBELS-M measures continued to be significant in both hierarchical analyses. Inspection of the adjusted squared multiple Rs shows that after controlling for variance in general ability (K-BIT) and phonemic awareness (TOPA), the addition of the DIBELS-M explained a significant proportion of the remaining variance in students' scores on the three achievement measures. In fact, the hierarchical analyses show that the proportion of the variance explained by the DIBELS-M measures remained almost constant and, in the case of the WJ-R Skills Cluster and the Teachers' Rating Scale, increased. Furthermore, DIBELS-M scores explained a larger portion of the variance in achievement scores than did either the K-BIT or TOPA scores.

Across all analyses, LNF was the single best predictor of kindergarten achievement scores on the Broad Reading and Skills clusters of the WJ-R and of teachers' ratings of children's reading status. SNF had significant beta weights for the WJ-R Skills Cluster and the TOPA. PSA was a consistent significant predictor of teachers' ratings of children's reading status. IPA had a negative beta weight as a predictor of scores on the Teachers' Rating Scale.

The results of the simultaneous and the Design 1 hierarchical regression analysis us-

ing the Fluency and Ability Composite scores as predictors are presented in Table 8. The Fluency Composite score was a significant predictor of all criterion achievement measures. The partial correlation coefficients between the Fluency Composite and the achievement measures after the ability had been partialed out were all significant (.56 for WJ-R Broad Reading Cluster; .70 for the WJ-R Skill Cluster; .52 for the TOPA; and .56 for the Teacher Rating Scale.) The Ability Composite score was a significant predictor of TOPA scores.

Discussion

The present study represents an extension of the previous work (Good et al., 1992, 1996) evaluating the DIBELS measures by including a larger, more diverse nationally representative sample of kindergarten children. The results are consistent with the previous research on the DIBELS as well as with a large body of research on kindergarten level preliteracy abilities that have been associated with later reading acquisition (Blachman, 1984, 1989; Felton & Wood, 1989; Stahl & Murray, 1994). In addition, the strong correlations found between the DIBELS-M and the Woodcock-Johnson Skills Cluster confirm earlier findings (Daly, Wright, Kelly & Martens, 1997) of the relationship between prereading and math fluency in young children.

The data from this partial replication provide corroborative support for both LNF and PSA as well as preliminary support for an experimental measure, SNF. IPA emerged as the weakest measure evaluated in the study and may not have sufficient reliability and validity for individual student assessments, a conclusion also reached in the preliminary study by Good et al. (1992).

The concurrent validity estimates for the DIBELS-M LNF and PSA measures are generally weaker than correlations with the set of criterion measures used in previous studies by Good et al. (1992) and Kaminski and Good (1996). Although LNF and PSA emerged in Good and Kaminski's work as the strongest DIBELS measures, LNF and the experimental measure, SNF, yielded the highest reliability and concurrent validity coefficients in the present study.

The weaker correlations for both Ability measures (i.e., IPA and PSA) in the present study indicate the need for additional work on the instrumentation of these measures and improved training and administration procedures. For example, when multiple examiners are used to screen large groups of children, it may be beneficial to give examiners practice in making rapid scoring decisions. Practice is particularly important for the phonological awareness measures because the assessment of phonological analysis is a novel activity for many school staff members, including school psychologists. Because these 10-item measures are very brief (i.e., about 30 to 60 seconds), examiners are forced to make quick decisions regarding the accuracy of children's segmentation responses.

For school psychologists, the DIBELS measures represent many of the best features of alternative assessments. The results of the present study support the use of a subset of DIBELS-M measures by school psychologists: (a) to identify kindergarten children who would benefit from more intensive instruction, (b) to monitor the progress of these children in the acquisition of preliteracy skills, and (c) to evaluate the effectiveness of early prereading instruction (Shinn & Hubbard, 1992).

The combining of the four DIBELS-M measures into a battery format provides a more standardized assessment procedure, and the composite scores had stronger psychometric properties than did the individual DIBELS measures. In particular, the strong technical characteristics of the Fluency composite in the present study corroborates findings regarding CBM reading fluency for older students (Shinn, 1989). In addition, these composites allow further comparisons between the performance of individual children as well as between groups of children.

The DIBELS measures are practical because they are very brief, easily repeated, and adapted to the curriculum. They do not require elaborate materials and can be readily administered by school psychologists and other school-based personnel with minimal training. The measures are easy to score, and children should benefit from exposure to these skills.

Feedback from the repeated assessments may encourage more kindergarten teachers to emphasize these preliteracy skills and to incorporate more opportunities for the development of phonological awareness and sound-symbol practice in the curriculum.

Although it would be desirable to assess preliteracy skills earlier in the kindergarten year in order to allow more time for interventions, the difficulty of these measures for kindergarten-aged children may restrict their use until the end of the year. These measures have also been examined for end-of-year first grade students but they may have greater utility for identifying beginning-of-year first graders who are at-risk for reading failure and in need of more intensive interventions targeting preliteracy skills. Kaminski and Good (1996) have reported preliminary data on the predictive validity of select DIBELS measures with end-of-year reading status in first grade. Additional studies of reliability and validity with larger populations of students are needed. Given the limitations of current screening and readiness instruments (Wenner, 1995), the DIBELS measures provide systematic assessment data for use in the early identification of children who are at-risk for reading failure.

References

Bagnato, S. J., Neisworth, J.T., & Munson, S. M (1989) *Linking developmental assessment and early intervention: Curriculum-based prescriptions* (2nd ed.). Rockville, MD: Aspen.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49-66.

Blachman, B A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76*, 610-622.

Blachman, B.A. (1989). Phonological awareness and word recognition: Assessment and intervention In A. B Kamhi & H. W. Catts (Eds.), *Reading disabilities: A developmental language perspective* (pp 133-158). Boston: College-Hill.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor, MI: University of Michigan Press.

Burns, J. M., Collins, M D., & Pausell. J. C (1991). A comparison of intellectually superior preschool accelerated readers and nonreaders Four years later. *Gifted Child Quarterly, 35*, 118-124.

Burstein, L (1994). *Performance-based assessment for accountability purposes: Taking the plunge and assessing the consequences*. National Center for Research on Evaluation. Standards and Student Testing. Los Angeles, CA. ED 3826760.

Clark, E. (1995). [Review of the Developing Skills Checklist] In J. C. Conoley & J. C. Impara (Eds), *The twelfth mental measurements yearbook* (pp. 278-281). Lincoln, NE Buros Institute of Mental Measurements.

CTB Macmillan/McGraw-Hill. (1990). *Developing Skills Checklist*. Riverside, CA. CTB/McGraw-Hill.

Cummings, J. A. (1994). [Review of the Woodcock-Johnson Psycho-Educational Battery-Revised.] In J. C. Conoley & J. C Impara (Eds.). *The twelfth mental measurements yearbook* (pp. 1113-1116). Lincoln, NE: Buros Institute of Mental Measurements.

Daly, E J. Wright, J A., Kelly, S. Q., & Martens, B. K. (1997) Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12*. 268-280.

Deno, S (1985). Curriculum-based measures: The emerging alternative *Exceptional Children. 52*, 219-232.

Deno, S., & Fuchs, L (1987) Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16

Drieling, L., & Copeland, E. (1988, August). *A model for the evaluation of kindergarten screening tests*. Paper presented at the annual meeting of the American Psychological Association. Atlanta.

Dunn, C. (1992). A validation study of the Detroit Tests of Learning Aptitude-Primary (2nd ed.). *Diagnostique, 17*, 266-272

Ehri, L. C., & Wilce, L S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20*, 163-179

Felton, R. H., & Wood. F.B (1989) Cognitive deficits in reading disability and attention deficit disorder. *Journal of Learning Disabilities, 22*, 3-13.

Fletcher, J. M., Shaywitz, S E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K K., Francis, D J., Fowler, A. E., & Shaywitz, B A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6-23

Fuchs, L. S. (1994). Integrating curriculum-based measures with instructional planning for students with learning disabilities. In N C. Jordan & J. Goldsmith-Phillips (Eds), *Learning disabilities. New directions for assessment and intervention* (pp. 166-196). Boston: Allyn & Bacon

Fuchs, L S., & Deno, S L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15-24.

Fuchs, L S., Deno, S., & Mirkin, P (1984). Effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning *American Education Research Journal, 21*, 449-460.

Fuchs, L. S., & Fuchs, D (1997) Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30*, 1-16.

Germann, G., & Tindal, G. (1985). An application of curriculum-based assessment: The use of direct and repeated measurement. *Exceptional Children, 52,* 244-265.

Good, R. H. (1994, April). *New developments and extensions of CBM Dynamic indicators of basic early literacy skills.* Paper presented at the Council for Exceptional Children Annual Convention, Denver.

Good, R , Kaminski, R., Laimon, D., & Johnson, D. (1992). *Advances in assessment for the primary prevention of early academic problems* (Summary). The Oregon Conference Monograph. Eugene, OR: University of Oregon.

Herman, J. L., Aschbacher, P R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Jackson, N. E., Donaldson, G W., & Mills, J. R. (1993). Components of reading skills in post kindergarten precocious readers and level-matched second graders. *Journal of Reading Behavior, 25,* 181-206.

Kaminski, R A , & Good, R. H (1996) Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215-227.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test.* Circle Pines, MN: American Guidance Service, Inc.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program Findings and implications. *Educational Measurement: Issues & Practices 13,* 5-16.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (1992, April). *Empirical evidence for the reliability and validity of performance assessments.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Linn, R. L., & Baker, E L (1996) Can performance-based student assessments be psychometrically sound? In J. B Baron & D. P. Wolf (Eds.), *Performance-based student assessment Challenges and possibilities* (pp. 84-103). Chicago: The University of Chicago Press.

Marston, D., & Magnusson D (1985) Implementing curriculum-based measurement in special and regular education settings *Exceptional Children, 52,* 266-276.

Marston, D., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation In J. L. Graden, J E Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp 137-172). Washington, DC: National Association of School Psychologists.

Martelle, Y , & Smith, D. (1994,March) *Relationship of the WIAT and WJ-R tests of achievement in a sample of students referred for learning disabilities* Paper presented at the Annual Meeting of the National Association of School Psychologists, Seattle, WA.

McGrew, K S , Werder, J. K., & Woodcock, R. W (1991) *WJ-R technical manual.* Allen, TX: DLM Teaching Resources.

Meisels, S. J., Wiske, M. S., & Tivnan, T. (1984). Predicting school performance with the early screening inventory *Psychology in the Schools, 21,* 25-33

Miller, W H (1995). *Alternative assessment techniques for reading and writing* West Nyack, NY. The Center for Applied Research in Education.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62,* 229-258.

Prewett, P. N. (1992). The relationship between the Kaufman Brief Intelligence Test (K-BIT) and the WISC-R with referred students. *Psychology in the Schools, 29,* 25-27.

Prewett, P. N., & McCaffery, L. K. (1993). A comparison of the Kaufman Brief Intelligence Test (K-BIT) with the Stanford-Binet, a two-subtest short form, and the Kaufman Test of Educational Achievement (K-TEA) brief form. *Psychology in the Schools, 30,* 299-304.

Share, D. L., Jorm, A. F., MacLean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76,* 1309-1324.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21,* 22-27.

Shinn, M. R. (1989). *Curriculum-based measurement.* New York: Guilford.

Shinn, M R. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 547-567). Washington, DC: National Association of School Psychologists.

Shinn, M. R , Good, R. H., Knutson, N., & Tilly, D. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459-479.

Shinn, M. R., & Hubbard, D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. *Focus on Exceptional Children, 24*(5), 1-20.

Shull-Senn, S., Weatherly, M., Morgan, S. K., & Bradley-Johnson, S. (1995) Stability reliability for elementary-age students on the Woodcock-Johnson Psychoeducational Battery-Revised (Achievement Section) and the Kaufman Test of Educational Achievement. *Psychology in the Schools, 32,* 86-92.

Smith, D K (1992, August) *Relationship of the K-BIT and SIT-R in a gifted sample.* Paper presented at the annual meeting of the American Psychological Association. Washington, DC.

Stahl, S A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86,* 221-234.

Stevenson, H W , Parker, T., Wilkinson, A., Hegion, A., & Fish, E (1976) Longitudinal study of individual differences in cognitive development and scholastic achievement. *Journal of Educational Psychology, 68,* 377-400.

Stipek, D., Feiler, R , Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development, 86,* 209-223.

Stoner, G (1992) Validating curriculum-based measurement Essential concerns from a behavioral perspective *School Psychology Quarterly, 7,* 17-21.

Tindal, G (1993) A review of curriculum-based procedures on nine assessment components. In J. J. Kramer

(Ed.), *Curriculum-based measurement* (pp. 25-64). Lincoln, NE: Buros Institute of Mental Measurements.

Torgesen, J. K., & Bryant, B.R. (1994). *Test of Phonological Awareness*. Burlingame, CA. Psychological and Educational Publications, Inc.

Torgesen, J. K., Morgan, S. T., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84*, 364-370.

Tramontana, M. G., Hooper, S. R., & Selzer, S. C (1988). Research on the preschool prediction of later academic achievement: A review. *Developmental Review, 8*, 89-146.

Vellutino, F. R., & Scanlon, D. M. (1987) Phonological coding, phonological awareness, and reading ability: Evidence from longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321-363.

Wenner, G. (1995). Kindergarten screens as tools for the early identification of children at-risk for remediation or grade retention. *Psychology in the Schools, 32*, 249-254.

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25*, 9-23.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.

Jacquelyn Elliott, Ph.D., is a practicing school psychologist and has worked in the Topeka Public Schools, Shawnee Heights Schools, and for the Menninger Center for Learning Disabilities in Topeka, Kansas. She completed her doctorate at the University of Kansas. Her professional interests include reading disabilities and reading acquisition in young children.

Steven Lee received his Ph.D. from the University of Nebraska-Lincoln in School Psychology in 1986. He has been on the faculty of the University of Kansas for 12 years. He is the Director of the School Psychology Training Program. He teaches courses in psychological consultation, classroom management and field experiences in school psychology. His research interests are in self-monitoring, inclusionary practices, classroom observation systems, student assistance teams, school crises/violence and classroom management. He is a principal investigator at the School Equity and Excellence Model (SEEM) Center and manages a School Level Program Evaluation Grant funded by the Kauffman Foundation in the Kansas City, Kansas Public Schools.

Nona Tollefson, Ph.D., is a professor of educational psychology and research at the University of Kansas. She teaches courses in assessment and research methodology and offers staff development programs in the areas of assessment and program evaluation.