AN EXAMINATION OF THE RELIABILITY AND VALIDITY OF DYNAMIC INDICATORS OF BASIC SKILLS IN EARLY LITERACY ADMINISTERED IN KINDERGARTEN

by

WILLIAM CLAY CROWDER

(Under the Direction of Shanna Hagan-Burke)

ABSTRACT

The purpose of this study was to evaluate the reliability and validity of a series of dynamic indicators of basic skills in early literacy designed to identify students in need of early literacy intervention. A group of 177 kindergarten students were administered DIBS in early literacy at three points during the academic year: Four measures from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were administered along with AimsWeb Letter-Sound Fluency (LSF). The DIBELS measures included Initial Sound Fluency (ISF), Letter-Naming Fluency (LNF), Phonemic Segmentation Fluency (PSF), and Nonsense Word Fluency (NWF). Predictive criterion-related and concurrent criterion related validity were evaluated using correlation and multiple regression. Evidence supportive of validity was found for all measures, however measures involving skills with print were more strongly predictive of emergent reading than measures involving phonological awareness. In the fall of kindergarten, LNF was the strongest predictor of emergent reading at the end of kindergarten. At the middle and end of kindergarten, NWF was the strongest predictor of emergent reading. The results of this study are supportive of using fluency-based measurements of component literacy skills, specifically skills
with letter-naming (LNF) and letter-sound identification (NWF and LSF), to identify students in need of early academic intervention in emergent literacy skills.

INDEX WORDS: Reliability, Validity, Assessment, Formative assessment, Early literacy, Phonological awareness, Reading acquisition, Kindergarten, Dynamic indicator, curriculum-Based Measurement, Fluency
AN EXAMINATION OF THE RELIABILITY AND VALIDITY OF DYNAMIC INDICATORS OF BASIC SKILLS IN EARLY LITERACY ADMINISTERED IN KINDERGARTEN

by

WILLIAM CLAY CROWDER

B.A., New York University, 1992

M.Ed. The University of Georgia, 2000

A Dissertation Submitted to the Graduate Faculty of The University Of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005
AN EXAMINATION OF THE RELIABILITY AND VALIDITY OF DYNAMIC INDICATORS OF BASIC SKILLS IN EARLY LITERACY ADMINISTERED IN KINDERGARTEN

by

WILLIAM CLAY CROWDER

Major Professor: Shanna Hagan-Burke
Committee: Mack D. Burke
Noel Gregg
Cynthia Vail
Roy Martin

Approval for Electronic Submission:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2005
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................ v

CHAPTER

1 INTRODUCTION .....................................................................................................................1

2 LITERATURE REVIEW .........................................................................................................8

3 METHOD ...............................................................................................................................46

4 RESULTS ..............................................................................................................................57

5 DISCUSSION .......................................................................................................................68

REFERENCES ............................................................................................................................85

APPENDICES ..........................................................................................................................104

A DIBELS INITIAL SOUND FLUENCY ..................................................................................104

B DIBELS PHONEMIC SEGMENTATION FLUENCY ..............................................................106

C DIBELS LETTER-NAMING FLUENCY ...............................................................................108

D DIBELS NONSENSE WORD FLUENCY ...........................................................................110

E AIMSWEB LETTER-SOUND FLUENCY .............................................................................112

F DIBELS ORAL READING FLUENCY ....................................................................................114
LIST OF TABLES

Table 1: Summary of Reliability Estimates for DIBS in Early Literacy ........................................116
Table 2: Summary of Validity Evidence for DIBS in Early Literacy ...............................................117
Table 3: Summary of DIBELS Cut-Scores and Norm-Group Mean Scores for A-LSF ................119
Table 4: Participant Characteristics ...............................................................................................120
Table 5: Schedule of Test Administrations for Kindergarten .......................................................121
Table 6: Reliability Estimates for DIBS and Spring Criterion Measures ........................................122
Table 7: Descriptive Statistics .......................................................................................................123
Table 8: Correlation Matrix of DIBS and Criterion Measures of Reading .....................................124
Table 9: Correlations Among DIBS from Fall, Winter, and Spring Benchmarks with TERA-3 Subtests ........................................................................................................................................125
Table 10: Predicting D-ORF and TERA-3 from Fall DIBS: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression .................................................................126
Table 11: Predicting D-ORF and TERA-3 from Winter DIBELS and A-LSF: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression ........................................127
Table 12: Concurrent Validity of Spring DIBELS and A-LSF with D-ORF and TERA-3: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression .........................129
CHAPTER 1
Introduction

Foremost among the goals of the No Child Left Behind Act (NCLB, U. S. Department of Education, 2002) is ensuring that all students read competently by the end of the third grade. Students failing to meet state standards of reading performance based on standardized achievement tests are provided with remedial services or retained. However, a students’ performance must fall far enough behind that of his peers that a difference in their relative achievement is measurable with a norm-referenced test. The problem lies not with the test, but rather the timing: By as early as the end of first grade, many critical opportunities to provide intervention to struggling readers have passed. Indeed, most students with poor reading skills in the early grades are not likely to improve later (Cunningham & Stanovich, 1998; Juel, 1988; Stanovich, 1986). In light of this evidence, preventing reading failure as early as possible is the most sensible approach to improving reading outcomes for students (Adams, 1990; Good, Simmons, & Kame’enui, 2001). Prevention-oriented approaches require that students at risk of reading failure be identified as early as possible and their progress in critical early literacy skills be monitored closely. For both of these tasks, most summative norm-referenced achievement tests are inappropriate. The use of dynamic indicators of basic skills (DIBS) in early literacy represents one approach to measurement with the technical characteristics necessary for identifying students at-risk of reading failure and for monitoring students’ growth in response to early literacy instruction (Good & Kaminski, 1996; Good & Kaminski, 2002; Shinn, 1995). The present study examines the reliability and validity of a series of DIBS in early literacy used
kindergarten to identify struggling readers and to predict outcomes on year-end measures of reading ability.

DIBS were developed based on Curriculum-Based Measurement (CBM, Deno, 1985). CBM is an alternative form of assessment originally developed for use within a problem-solving model of addressing academic achievement problems. A substantial body of research supports the technical adequacy of CBM (Deno, 1989; Deno & Fuchs, 1987; Fuchs et al., 1984; Fuchs & Fuchs, 1997; Germann & Tindal, 1985; Marston & Magnusson, 1985; Shinn, 1989; Shinn & Hubbard, 1992). CBM probes are short, fluency-based measures derived from the curriculum. Administration takes only a few minutes and can be carried out by classroom teachers. CBM probes are administered frequently and, unlike many norm-referenced tests, are sensitive to small changes in performance across relatively short periods (Deno, 1985; Fuchs, 1989; Fuchs et al, 1984; Fuchs & Fuchs, 1991, 1999; Good & Shinn, 1992; Stecker & Fuchs, 2000). CBM has been used effectively to identify students performing below their same-age peers (Elliot & Fuchs, 1997; Marston & Magnusson, 1985; Marston et al, 1984; Shinn & Marston, 1985), and has been used on system-wide scales to formatively monitor the academic achievement of all students. These applications of CBM have been found more efficient and effective than assessment systems relying on summative norm-referenced approaches to measurement (Germann & Tindal, 1985; Greenwood, Tapia, Abbot, & Walton, 2003; Howe, Scierka, Gibbons, & Silberglitt, 2003; Ikeda, Tilly, Stumme, Volmer, & Allison, 1996; Simmons, Kuykendall, King, Cornachione, & Kame’enui, 2000; Tilley & Grimes, 1989).

Based on the strong research support for the utility and technical adequacy of CBM, Shinn (1995) suggested that the use of CBM could be extended to earlier academic contexts by conceptualizing CBM more broadly as a set of dynamic indicators of basic skills or DIBS. Both
CBM and DIBS draw on similar measurement logic—short, fluency-based probes can produce valid indications of overall performance, but in a more efficient fashion that most traditional achievement tests (Shinn, 1995). CBM probes are typically derived from the curriculum, with the identification and remediation of existing achievement problems as the goal. DIBS in early literacy measure competence with foundational reading skills that correlate to later reading ability. By identifying students at-risk for failure early, academic intervention can be provided before the critical window of successful reading acquisition begins to close and the benefits of interventions diminish (Good et al., 1998; Good et al., 2001). In the present study, “DIBS” was used generically to refer to the series of fluency-based measures of early literacy skills that was evaluated.

The *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) is a set of published measures that function as DIBS in early literacy. DIBELS were developed by researchers at the University of Oregon as formative measures of early literacy skills for use within a preventative approach to reading failure (Good, Kaminski, Shinn, Simmons, Kame’enui, & Wallin, 2003). Four DIBELS measures administered in kindergarten were examined in this study (see methods section for a more thorough description of DIBELS measures). Each is designed to measure important basic skills related to reading acquisition: phonological and phonemic awareness, alphabetic understanding, and fluent word recognition in connected text (Adams, 1990; Blachman, 2000; Ehri, Nunes, Stahl, & Willows, 2001; Good, et al., 1998; Good, et al., 2001; Kame’enui & Simmons, 2001; Kaminski & Good, 1996; Simmons & Kame’enui, 1998; National Reading Panel, 2000; National Research Council, 1998; Torgesen, Wagner, & Rashotte, 1997). The technical characteristics of DIBELS measures are similar to those of CBM, making DIBELS well suited for screening, benchmark assessment, frequent progress monitoring, and instructional
decision-making (Good et al., 2001). Previous studies evaluating the technical characteristics of DIBELS and other DIBS of early literacy skills suggest that most are reliable and valid for identifying students at-risk for reading failure and formatively assessing student growth (Burke, Hagan-Burke, & Ditkowsky, in review; Crowder, Burke, & Hagan-Burke, in review; Elliot, Lee, & Tollefson, 2003; Good, Gruba, & Kaminski, 2001; Good & Jefferson, 1998; Good, Kaminski, Laimon, & Johnston, 2000; Good, et al., 2001; Good, et al., 2003; Hagan-Burke, Burke, & Crowder, in press; Hintze, Ryan, & Stoner, 2003 ). However, these studies alone are insufficient evidence. If these tools are to be adopted for use within current standards-based accountability systems and replace summative achievement tests as the default standard, more empirical support is needed. Studies supporting the validity of DIBS targeting early literacy skills are needed that replicate previous findings and provide new and potentially stronger support for the validity of existing measures. More comprehensive evaluations of validity are also needed that move beyond the frequently reported estimations of criterion-related and predictive validity and towards more multi-faceted examinations of construct validity that include an examination of the social consequences of using these measures to make decisions.

The purpose of this study was to evaluate the validity of a series of DIBS designed to target critical early literacy skills in kindergarten. Accordingly, *Letter-Sound Fluency (A-LSF)* (Shinn & Shinn, 2002) was administered concurrently with DIBELS at each scheduled administration (fall, winter, and spring). DIBELS lacks a direct measure of letter-sound knowledge independent of sound blending. The following DIBELS measures and an additional measure of letter-sound knowledge were evaluated:

*DIBELS-Initial Sound Fluency (D-ISF)* (Good & Kaminski, 2002)

*DIBELS Letter Naming Fluency (D-LNF)* (Good & Kaminski, 2002)
DIBELS Phoneme Segmentation Fluency (D-PSF) (Good & Kaminski, 2002)

DIBELS Nonsense Word Fluency (D-NWF) (Good & Kaminski, 2002)

The addition of A-LSF creates a broader range of indicator skills than DIBELS measures alone. Previous studies have suggested that knowledge of letter-sounds could be an effective predictor of reading, although this is frequently overlooked in favor of letter-naming ability, which has the strongest empirical support as a predictor of reading (Elliot, Lee, & Tolefson, 2001; Shinn & Shinn, 2002; Speece & Case, 2001).

In addition to DIBELS and A-LSF, the Test of Early Reading Ability-Third Edition (TERA-3) (Reid, Hresko, & Hammill, 2001), was administered in the spring of the academic year and used as a criterion measure of reading. The TERA-3 is a published norm-referenced test with established reliability and validity, measuring early literacy skills within three subtest domains, Alphabet, Conventions, and Meaning. As a second criterion measure, DIBELS Oral Reading Fluency (D-ORF, see Good & Kaminski, 2002) was administered in the spring. Although D-ORF is not typically administered until the middle of the first grade, several studies have suggested the importance of using measures in kindergarten that reflect the act of reading (i.e., reading connected text) in addition to measuring component skills like phonemic awareness and letter-sound knowledge (Speece, et al., 2001; O’Connor, 2000).

The following research questions regarding the reliability, concurrent criterion-related validity, and predictive criterion-related validity of the set of DIBS examined in this study were posed:
Reliability

Question 1. Will scores from identical forms of DIBS (DIBELS subtests and A-LSF) administered two weeks apart correlate strongly \((r>.80)\), demonstrating that the skills measured are stable?

Predictive Criterion-Related Validity

Question 2. What are the correlations among DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF), middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF), and end (D-LNF, D-PSF, D-NWF, and A-LSF) of kindergarten?

Question 3. What are the correlations between DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF) and middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF) of kindergarten, and criterion measures of reading ability administered in the spring (TERA-3, D-ORF)?

Question 4. Of those administered at the beginning and middle of kindergarten, will particular DIBELS measures explain a unique amount of the variance of criterion measures of reading administered at the end of kindergarten (TERA-3 and D-ORF)?

Question 5. Will A-LSF, administered at the beginning and middle of kindergarten account for a significant amount of variance in criterion measures of reading (TERA-3 and ORF) after controlling for the effects of DIBELS?

Concurrent Criterion-Related Validity

Question 6. What are the correlations between DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) and criterion measures of reading (TERA-3 and ORF)?
Question 7. Will certain DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) explain unique amounts of variance of criterion measures of reading (TERA-3 and D-ORF)?

The researcher analyzed a set of extant data consisting of scores for 177 kindergarten students who were formatively assessed using the DIBS in early literacy and the criterion measures of this study. Two sets of analyses were conducted that examined validity from multiple perspectives. Correlation and hierarchical regression were used to evaluate reliability, predictive validity, and concurrent criterion-related validity of the DIBS evaluated in the study.

Correlation coefficients were computed as estimates of test-retest reliability and inter-rater reliability. Simultaneous and hierarchical regression analyses were conducted to evaluate the predictive validity of each measure administered in the fall and winter of kindergarten using the TERA-3 and D-ORF as outcome measures of emergent reading ability. Similar analyses were conducted to evaluate the concurrent criterion-related validity of the literacy measures administered in the spring with the TERA-3 and D-ORF. The amount of variance in each criterion measure explained by the individual DIBS was calculated.

The results of this study contribute to a growing body of research supporting the use of DIBS in early literacy as valid, formative, assessment tools focused on preventative approaches to reading failure. Furthermore, the results help build the foundation of technical evidence necessary for future studies to evaluate DIBS in early literacy from a treatment validity perspective (Heller, Holtzman, & Messick, 1982; Messick, 1989).
CHAPTER 2

Literature Review

Introduction

The purpose of this review is to summarize the relevant research literature regarding the development, application, and technical adequacy of dynamic indicators of basic skills (DIBS) in early literacy. Of particular interest was the use of those measures to identify students at-risk for reading failure and monitor the progress of students’ reading acquisition during the kindergarten year. The first section of this review provides some context for a discussion of the DIBS in early literacy examined in the study and includes (a) a brief overview of the development of CBM, from which DIBS in early literacy were developed; (b) an explanation of the measurement logic that distinguishes DIBS from traditional norm-referenced measurement, and (c) an explanation of the technical similarities and differences between DIBS and CBM.

The second section of the review begins with a description of early literacy skills that correlate with later reading and that are measured by the DIBS in early literacy evaluated in this study. Research investigating the relations among these skills and reading acquisition is reviewed and dominant theories of early reading acquisition are discussed. The use of DIBS to measure early literacy skills and for monitoring student progress are discussed. Previous studies evaluating the technical adequacy of DIBS in early literacy, primarily DIBELS, are reviewed. The need for further research, particularly studies designed to examine the validity of DIBS in early literacy used in kindergarten is established, and research questions pertaining to reliability, predictive criterion-related validity, and concurrent criterion-related validity are framed.
Curriculum-Based Measurement: The Foundation for Dynamic Indicators of Basic Skills

Over three decades ago, Dr. Stanley Deno of the University of Minnesota saw the need for measurement tools that his student teachers could use formatively to measure special education students’ growth and evaluate the effectiveness of instruction (Deno, 1970). The summative assessment data that schools collected from traditional norm-referenced tests of achievement were not suited for monitoring progress (Cohen, 1976). Drawing on research from the field of applied behavior analysis, Deno set out to design a set of measurement tools with technical characteristics that could be used for frequent measurement, would result in data that could be graphed and analyzed visually, and would be sensitive to small changes in performance over a short time (Deno, 1985, Deno, 2003a, Shinn, 1995). In 1977, after several years of field-testing, Data-Based Program Modification (DBPM) was developed (Deno & Mirkin, 1977). The DBPM model outlined how data gathered from short formative assessment measures could be used for evaluation and program modification decisions for students in special education (Deno, 2003a). The measures within the model, although intuitively appealing, lacked empirical evidence to support their use as formative evaluation tools. Over the next few years, Deno and his colleagues examined these and other early formative evaluation measures with a particular interest in their technical characteristics (e.g., reliability and concurrent validity), and their feasibility for use within classrooms on a frequent basis (Shinn, 1995). Formative evaluation measures were required to (a) be tied to the curriculum; (b) be technically adequate, demonstrating sufficient reliability and validity; (c) be standardized, using procedures, materials, administration, and scoring that adhere to prescribed guidelines; (d) be logistically feasible (e.g., multiple forms, time-efficient, easy to teach and understand); (e) be sensitive to change, thus having the capacity to model growth over a short time (Deno & Fuchs, 1987). Studies between
1978 and the mid 1980s provided support for the technical adequacy of measures of reading, particularly oral reading fluency (ORF), originally a component of DBPM. New measures were also developed for formatively assessing reading, spelling, and written expression (e.g., Deno, Mirkin, & Chiang, 1982; Deno, Marston, & Mirkin, 1982; Deno, Mirkin, Lowry, & Kuehnle, 1980; Fuchs, Tindal, & Deno, 1984).

Until the 1980s, the system of simple formative measurements that had been developed by Deno and his associates (that later became known as CBM) were considered to be component of a larger approach to assessment called Curriculum-Based Assessment (CBA) (Tucker, 1985). At the time, a movement advocating the grounding of measurement of student progress in the curriculum had gained considerable momentum. This was in response to the use of the widespread use of PNRTs and diagnostic assessments that was driving educational practice at the time (Shinn, 1998). However, CBA was a very broadly defined concept that, although including curriculum-derived materials and repeated measurement, lacked the qualities necessary to be suitable for valid formative evaluation. Most notably, there was little research to support the technical adequacy of CBA assessments. The term Curriculum-Based Measurement (CBM) was formally adopted in 1985 and served to distinguish the formative evaluation tools Deno and his colleagues had developed from other informal assessments that were subsumed under the broad category of CBA (Deno, 1985). In addition to having a strong research base to support its use, the intended purpose of CBM measurements require this distinction to be made. CBA was used (and continues to be used) for informal measurement of short-term progress or for identifying whether students have mastered certain aspects of curriculum. CBM, on the other hand, is designed for monitoring student progress over time and evaluating the effectiveness of interventions. These purposes require characteristics that CBA approaches do not typically
incorporate. CBA uses the actual curriculum materials that are taught for the purposes of the assessment. This results in a very reliable and valid assessment of whether the student has mastered a particular set of skills or learned particular information, but does not reveal how well he or she is progressing in the curriculum in general, because the assessment materials (and essentially the construct measured) are continually changing (Deno, 1985). By design, CBM requires that test content be related to the curriculum, but also requires that the content across tests be of equivalent difficulty (i.e., standardized). With the difficulty held constant, CBM measures can index a student’s progress in relation to the overall curriculum that is represented by the equivalent measures (Deno, 1986).

Requiring that the test materials be of equivalent difficulty gives CBM a degree of utility that CBA does not have. Deno (1985) offered the metaphor of using a ruler to measure a child’s height. To measure physical growth, one uses a consistent standard of reference for comparison (the ruler). Changes in the child’s height are evident (and trusted as accurate) because earlier measurements, when compared to current measurements, are referenced against the same metric (the ruler). If a different standard of measurement were chosen each time height were measured (e.g., using the length of a candy bar instead of a ruler) comparisons between points in time for the same individual would be virtually meaningless. Thus, standardizing content difficulty maintains consistency across multiple measurements, and this allows CBM to model growth accurately.

In addition to allowing comparisons within an individual across time, standardizing the measurement tools allows comparisons between individuals as well (Deno, 1986). Relying again on the height metaphor, because rulers are standardized tools (i.e., they are consistently the same), comparisons of height measurements among children can be made if they are measured
with similar rulers. For example, the initial differences in height between children may be measured or the differences in their rates of growth over time through multiple measurements. The effectiveness of a hypothetical medical intervention intended to increase the height of an individual (or group of individuals) could be evaluated using the same data. Likewise, CBM can be used to make initial comparisons between individuals, compare rates academic of growth over time, evaluate the effectiveness of programs, and in other varied decision-making contexts described later (Deno, 2003b).

Another important distinction between CBA and CBM is the selection of the assessment materials. CBA often takes the form of one of two major measurement models, mastery measurement and general outcome measurement. Mastery measurement complements common (but not necessarily effective) teaching practices fairly well (Fuchs & Deno, 1991). Within mastery measurement, global skills are broken down into component skills that are translated into short-term academic objectives. These objectives are taught, mastered, and assessed, often using the same content for assessment as was used for instruction. This practice provides a fair estimate of short-term progress, but seriously limits making inferences beyond that. The second model, general outcome measurement, focuses on the measurement of longer-term goals (e.g., expected level of reading proficiency at year-end) (Fuchs & Deno, 1994). This requires that the materials used for assessment represent the curriculum content and difficulty at the time of the goal, but it is not necessary that materials be pulled directly from the curriculum (Marston, 1989). In this model, the assessment data do not represent performance on a specific section of the curriculum. Instead, data represent more general indications of broad performance within the curriculum as a whole (Fuchs & Deno, 1994). Unlike the content from PNRTs, CBM materials can be both standardized and meaningfully connected to the curriculum (Deno, 2003b). What
results is not a direct measurement of curriculum mastery, but rather an indication of performance broadly within the curriculum.

Scores on CBM assessments provide a general estimate of a broad range of skills within a domain, but are not intended to measure all skills within a domain. Essentially, they are intended to measure critical skills (e.g., reading fluency) that serve as barometers of overall competence (Shinn, 1989). This concept has been one of the barriers to the acceptance of CBM because educators are not accustomed to drawing inferences about overall achievement from measurements of isolated skills used as indicators. Medical thermometers are a good example of an indicator as they can reveal meaningful information without diagnosing a specific problem: They can indicate that a problem exists (fever), they can indicate the severity of the problem (fever of 104 degrees), they can indicate whether an intervention is working (fever is reduced after medication), and indicate when illness subsides (fever returns to normal) (Shinn, 1998). What they cannot do, however, is reveal exactly what the nature of the sickness is, only that a problem exists. Shinn (1998) encourages thinking of CBM as “educational thermometers”: CBM measures cannot diagnose the nature an academic problem with precision but can reliably detect the presence of problems, indicating the need for more precise types of diagnostic assessment.

*Description of CBM Measures*

Because CBM measures are designed for formative assessment, all share certain characteristics regarding construction, administration, and scoring. The characteristic that most clearly distinguishes CBM measures and DIBS from traditional assessments is that CBM and DIBS are fluency-based: The scores generated by the measurements are based on the number of correct items within a set amount of time. CBM measures are brief, usually requiring only
several minutes for administration and all measures are intended to have multiple forms used for monitoring growth through repeated measurement. CBM measures are also effective at differentiating between proficient and non-proficient students. Most PNRTs have this kind of sensitivity; however, scores tend to reflect only longer-term cumulative changes in performance, making them ineffective for modeling short-term growth. CBM measures, on the other hand, are designed to reveal more subtle changes in performance that indicate a student’s response to instruction (Marston, 1989). Studies have shown that CBM measures are sensitive to individuals’ growth over the short-term (Fuchs, Fuchs, & Hamlet, 1990; Fuchs, Fuchs, & Fernstrom, 1993) and to initial differences between individuals (Shinn & Marston, 1985; Shinn, Yssledyke, Deno, & Tindal, 1986).

The majority of studies that support the use of CBM have focused on reading, and several factors account for this. Foremost, is that reading provides a student with access to the rest of the curriculum; reading can be thought of as a pre-requisite skill for nearly every other academic domain (Adams, 1990). If students do not learn to read proficiently in the early grades, there is a high degree of certainty that they will not do so later (Juel, 1988). Current measurement practices in schools typically relies on summative achievement testing to identify students with reading problems; however, norm-referenced tests cannot detect a problem until one has developed, usually in the second or third grade (Marston, 1989). Remediation and special education services are the options schools take, but the unfortunate truth is that neither remediation nor special education are particularly effective at correcting reading problems (Johnson & Allington, 1991; Vaughn, Moody, & Schumm, 1998). Early identification and prevention have been accepted as the most sensible solutions to problems with literacy (Adams, 1990; Torgesen, 1998, Good & Kaminski, 1996). Thus, research on the development and
validation of new measures used for preventing reading problems has taken on a new importance.

Concerns Regarding CBM Measures

Face validity is perhaps the most common concern educators have when considering the implementation of CBM within a formative evaluation system (Deno, 2003a). In fact, problems with face validity, particularly with ORF, have caused the most resistance to implementing CBM (Yell, Deno, & Marston, 1989). The connection between the CBM or DIBS, and the skills they measure is not as evident when compared to traditional achievement tests. The idea of fluency-based measures is difficult for some to accept, particularly if the logic that underlies fluency measurements is not understood (Kame’enui & Simmons, 1998; LaBerge & Samuels, 1974; Shinn & Good, 1992; Wesson, King, & Deno, 1984). Many teachers are reasonably skeptical of a measure like ORF that purports to assess reading proficiency by measuring the speed of students' reading for only one minute. However, the connection between oral reading fluency and overall reading competence has a strong theoretical foundation and considerable empirical support (Kame’enui & Simmons, 1998; Fuchs, Fuchs, Hosp & Jenkins, 2001).

Several theoretical models of reading assist in conceptualizing oral reading fluency as an indicator of overall reading competence. LaBerge and Samuels (1974) describe reading as a complex skill that is made up of component processes (e.g. decoding, word recognition). The complex skill of reading and comprehending requires considerable attention, but one’s capacity to attend is limited. If attention is devoted to the lower-level mechanical components of reading, attention cannot be dedicated to higher processes like comprehension. If the component skills can be executed automatically, more attention is available for comprehension. Some have challenged aspects of the automaticity model of reading (Stanovich & Stanovich, 1995), but still
agree on key theoretical points, namely that efficiency with basic skills frees up capacity for higher-level processes (Fuchs, et al., 2001). Thus, fluency is a logical pre-requisite to comprehension (Carnine & Silbert, 1987), and should serve as an indicator of comprehension (i.e., overall competence). Studies have shown that this logic generalizes to the measurement of other basic skills (Deno, 2003a; Good et al., 2001; Fuchs & Fuchs, 1986; Shinn & Bomanto, 1998).

Extended Applications of CBM

Studies have shown that using CBM can lead to a variety of positive outcomes in classrooms including enhanced teacher planning and instruction (Fuchs, Deno & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Stecker, 1991; Fuchs, Fuchs, Hamlett, & Allinder, 1991), development of effective instructional groups (Fuchs, Fuchs, Bishop, & Hamlett, 1992; Wesson, 1992; Wesson, Vierthaler, & Haubrich 1989), and accurate evaluation of students’ response to instructional intervention (Fuchs & Fuchs, 1999; Fuchs, Fuchs, & Hamlett, 1990; Fuchs, Fuchs, Mathes, & Simmons, 1997; Howell, 1986; Marston & Magnusson, 1985; McMaster, Fuchs, & Fuchs, 2002). Outside of the classroom, CBM has been used to predict scores on high-stakes tests (Crawford, Tindal, & Stieber, 2000; Nolet & McLaughlin, 1997), and has even been used effectively to support formative evaluation in adult literacy (Bean & Lane, 1990). Research has also demonstrated the utility of measures that are specialized extensions of CBM logic—dynamic indicators of basic skills (DIBS). DIBS are designed for measuring students’ performance and monitoring progress in the acquisition of basic academic skills, usually before a formal curriculum has been introduced.

CBM as DIBS: Preventing Reading Failure
DIBS have been developed for the assessment of early literacy skills that serve as indicators of later reading competence (e.g., letter knowledge, phonemic awareness, word identification). The developers of DIBS in early literacy drew upon the formative evaluation logic of CBM: Growth is measured over time using repeated measurements, and inferences are made regarding overall competence based on the measurement of isolated skills. The use of DIBS in early literacy also capitalizes on the technical features of CBM that allow measures to be “uncoupled” from the curriculum. In this way, inferences can be made about a student’s performance in a curriculum based on measurements of skills that are important indicators of later reading competence, but that are not necessarily tied to a standard curriculum (Deno, 2003b). This is of particular importance when measuring the skills of students in kindergarten who are not typically engaged in formal reading instruction (Good & Kaminski, 1996). Here the application of DIBS to early literacy is different from CBM. Whereas CBM for reading is designed to identify students with existing reading problems, students in kindergarten do not have reading problems because they are not expected to know how to read (Good et al., 2001). The goal of using DIBS in early literacy is not to identify reading problems, but to identify students who are at-risk for later reading problems, and to prevent those problems before they begin (Kaminski & Good, 1996).

Dynamic Indicators of Basic Skills Used Within a Preventative Approach to Early Reading Failure

Context

The use of summative end-of-year achievement tests to identify students who are experiencing reading problems is standard practice in most systems (Carnine, 2000). The problem with this approach that reading problems are only detectable after they have developed,
and often this is not until several years of formal instruction have passed—as late as the third grade (Good & Kaminski, 1996). By this time, students have missed critical opportunities for corrective instruction and have had considerable time to develop unproductive attitudes toward school and their own abilities. In the majority of cases, students who experience early reading failure do not recover from it (Juel, 1998; Kame‘enui, 1993; Stanovich, 1986). Adams (1990) notes that the likelihood of a student in kindergarten successfully learning to read in the first grade is directly related to what they already know—an unfortunate fact considering how many children experiencing early failure go unidentified. Undoubtedly, more positive outcomes for children will result if ensuring accountability for student learning focuses on early prevention instead of remediation (Good & Kaminski, 1996). An accumulating body of research has demonstrated that with early identification and effective instruction, students can acquire the critical literacy skills necessary to establish successful learning trajectories and avoid failure (Lundberg, Frost, & Peterson, 1988; VanDerHeyden, Witt, Nauquin, & Noell, 2001; Vellutino, Scanlon, & Tazman, 1998). This hopeful notion has led researchers to develop efficient and technically valid dynamic indicators of basic skills designed to identify students at-risk for early reading failure and to monitor growth across time (e.g., Fuchs, Fuchs, & Compton, 2004; Good & Kaminski, 1996; Hintze, et al., 2001; Shinn & Shinn, 2002; O’Connor, 2000; VanDerHeyden et al., 2001).

DIBS in early literacy are conceptualized as downward extensions of CBM: Similar to the way that CBM-oral reading fluency in school age students indicates levels of overall reading performance, measurements of pre-reading skills in younger children indicate the child’s trajectory of future reading competence. DIBS in early literacy are not intended to provide an exhaustive assessment of all reading skills, but rather serve as fast and efficient indicators of
foundational skills that are necessary at critical stages. The selected DIBS evaluated in this study represent measurements of each of the following skills: Phonological awareness and phonemic awareness skills; alphabetic understanding, including knowledge of letter-names, letter-sound correspondences, and sound-blending skills; and fluency (Kaminski & Good, 1996, Shinn & Shinn, 2002). Reviews of correlational studies have demonstrated moderate to strong correlations (predictive and concurrent) between these skills and reading, and all have been identified as important components of successful reading acquisition (Adams, 1990). In the following section, all of the skills mentioned are defined, and the relationships of these early literacy skills to reading acquisition are discussed in more detail. An additional construct related to early reading, conventions of print (Reid, et al., 2001), is also described. Conventions of print refers to a broad set of awareness skills related to the use of printed language, but not to the specific knowledge of letter names or sounds (National Research Council, 1998).

Constructs of Early Reading Measured by Experimental DIBS in Early Literacy

Phonological and Phonemic Awareness. Phonological awareness has been broadly defined as the ability to perceive and manipulate the sounds of spoken words (Goswami and Bryant, 1990). The National Research Council (1998) defined phonological awareness as a general appreciation of the sounds of speech distinct from their meaning. Despite the subtle differences, both of these definitions suggest that phonological awareness has to do with the conscious awareness of the sounds that comprise spoken language. This broad ability encompasses the awareness of onsets (the initial sound in a word), rimes (the terminal portion of a word that contains a vowel), syllables, and the smallest unit of speech, the phoneme (Castles and Coltheart, 2004). Phonological awareness is demonstrated through a variety of auditory tasks, all conducted in the absence of print (e.g., initial phoneme identification, phoneme
deletion, rhyming, phoneme blending, and phoneme segmentation). Phonemic awareness is a more sophisticated type of phonological awareness, which requires the more fine-grained ability of identifying and manipulating individual phonemes (National Research Council, 1998).

Empirical studies have demonstrated the strong connection between phonological awareness and reading acquisition (see Ehri et al., 2001; Lundberg et al., 1988; National Reading Council, 1998; Scarborough, 1998). Children who have greater knowledge of the constituent sounds of words tend to be better readers than those whose knowledge is less developed (Adams, 1990; Goswami & Bryant, 1990; Wagner & Torgesen, 1987). Proficient phonological awareness, measured in kindergarten, is also predictive of future reading achievement (Juel, 1991; Scarborough, 1989; Stanovich, 1986; Wagner et al., 1994). Conversely, the lack of phonological awareness skill has been shown to be predictive of reading failure (Adams, 1990; Ball & Blachman, 1991). However, the majority of research indicates that phonological awareness is more predictive of superior reading than of deficient reading (National Reading Panel, 2000; Scarborough, 1998; Wagner & Torgesen, 1987), and frequently over-predicts future reading difficulties in students that will eventually learn to read effectively (O’Connor & Jenkins, 1999). Phonological awareness can be taught to children, and improvements in phonological awareness, specifically phonemic awareness skills have been shown to influence reading performance positively (Ball & Blachman, 1991; Bradley & Bryant, 1985; Lundberg et al.1988). The National Reading Panel (NRP, 2000) conducted a meta-analysis of studies investigating the effects of phonemic awareness training on reading. The Panel found an average effect size of $d=0.97$, and according to Cohen’s (1988) guidelines for interpreting the $d$ statistic, a value of $d=.97$ reflects a “large” effect. Based on these results, the Panel concluded that phonemic awareness instruction in kindergarten is a strong and positive influence on later
reading skills. This conclusion was challenged by Hammill (2004), who noted that the use of the $d$ statistic for evaluating the effects of instructional may present a potentially inflated impression of intervention effects, particularly for students with disabilities (Ives, 2003). Hammill (2004) transformed the $d$ statistics from the NRP meta-analysis to correlation coefficients ($r$s), resulting in the originally large effect sizes of phonemic awareness training reflecting moderate correlations ($r = .40$). Scarborough, (1998) in a meta-analysis of 27 studies investigating the predictive power of phonological awareness and later reading difficulty, found an average correlation of .46, and this value is consistent with Hammil’s (2004) reinterpretation of reported effect size statistics.

The existence of a relationship between phonological awareness and reading is undisputed; however, the relevance of that relationship remains the subject of debate. An examination of the research literature regarding the role phonological awareness plays in the process of reading acquisition reveals a wide range of divergent and sometimes overlapping conclusions. From these conclusions, three broad perspectives emerge each based on a unique explanation of the relationship of phonological awareness and reading. The first and perhaps most widely endorsed perspective is that phonological awareness is directly and causally related to reading acquisition, and without the necessary prerequisite knowledge of the sound structure of language, learning to read a phonetic alphabet will be impossible (Bradley & Bryant, 1985; Goswami & Bryant, 1990; Torgesen, Wagner, Rashotte, Rose, Lindamood, Conway, & Garvan, 1999). Central importance is placed on phoneme-level awareness of speech sounds, since printed English is composed of symbols (graphemes) that represent phonemes. Without an awareness of the individual phonemes in spoken language, making the connection between graphemes within words and the phonemes of spoken language they represent will be extremely difficult (Adams,
1990; Perfetti, Beck, Bell, & Hughes, 1987; Torgesen et al., 1999). In addition to phoneme-level awareness, awareness of larger units of spoken language such as rimes may play an important part in the development of early word recognition. Frequently occurring patterns of letters are associated with commonly used units of speech (e.g., *ing*, *tion*) (Goswami & Bryant, 1990). Students can make analogies be unknown words containing familiar letter-sequences and known words containing the same letter-sequences. (Goswami, 1993; Goswami & Bryant, 1990)

The second broad perspective regarding the relationship of phonological awareness to reading exchanges the causal role of phonological awareness for a consequential one. In this way, the act of learning to read focuses a child’s awareness on the phonemic structure of language, an aspect that is normally obscured from children’s’ awareness because spoken language is seamless (Ehri & Soffer, 1999). Unless they are taught to do so, children have no reason to become aware of individual phonemes until they begin learning to read and spell (Christensen, 1997; Ehri, 1989). Thus, the correlations between phonological awareness and reading demonstrated repeatedly in the literature are may be explained by the influence of instruction on the development of phonological awareness (Castles & Coltheart, 2003; Hammill, 2004).

Resting on the middle ground, the third perspective explains the relationship between phonological awareness and reading as reciprocal: both have causal influences on another (Perfetti, Beck, Bell, & Hughes, 1987; Stanovich, 1986; Ehri, 1989, 1990). The influences of one to the other may vary depending of the stage of reading acquisition. Ehri (1989) indicated that as children learn letter-sound correspondences in the early stages, they tend to become more aware of the phonemic segments of printed words. This in turn, leads to greater generalized awareness of spoken language on the phoneme level (Ehri, 1989). Later, as children learn to
read and spell phonemically irregular words, the ability to segment phonemes helps children to learn and remember the patterns or groups of phonemes that are encountered across words (Ehri, 1995; Ehri, 1998; Ehri & Sweet, 1991). Common to each of the three perspectives described is the undisputed importance of acquiring an understanding of the symbolic relationship of print to speech—alphabetic understanding.

*Alphabetic Understanding.* Alphabetic understanding has been defined as, “the mapping of print to speech and the phonological recoding of letter strings into corresponding sounds and blending stored sounds into words” (Good, Simmons, & Kame’enui, 2001, p. 261). Alphabetic understanding is measured with a range of print-related tasks including those that measure letter-name knowledge, letter-sound knowledge, the ability to blend printed letters into strings of sounds, the ability to blend phonemes into words. Unlike phonological awareness skills, there is less disagreement regarding the importance of alphabetic skills to reading acquisition:

Alphabetic understanding and the skills that result are integral to the act of reading (arguably, phonemic awareness skills are not) (Adams, 1990; Ehri, et al., 2001).

Measures of letter-naming ability have received a great deal of attention in predictive studies, consistently demonstrating strong correlations to later reading achievement.

Scarborough, (1998), in a review of studies that used letter-name knowledge as a predictor of later reading, found an average correlation coefficient of \(r = .52\). Kaminski and Good (1996) correlated results from letter-naming fluency tasks to the Metropolitan Readiness Test and found a coefficient of \(r = .77\). VanDerHeyden, et al. (2001), investigated the concurrent validity of letter-identification task and found significant and strong correlations to measures of phonological awareness \(r = .72\), letter-sound knowledge \(r = .76\), and, surprisingly, the composite math score of the Comprehensive Inventory of Basic Skills \(r = .67\) (Brigance, 1999).
Despite the strong predictive power of letter-name knowledge, teaching letter names would likely have little effect on reading achievement since letter naming is not required in order to read. Nor, for that matter, would teaching letter names increase math achievement, despite the strong relationship to math skills the previous study demonstrated. The relationship of letter-name knowledge to reading likely results from the similarity to letter-sound knowledge, another aspect of alphabetic understanding that is predictive of later reading (Speece et al., 2003; Vellutino et al., 1996). Ehri (1998) refered to letter-sound knowledge as grapheme-phoneme knowledge, which encompasses the understanding of regular as well as irregular phoneme-grapheme relationships that are present in many printed English words. The ability to understand one to one letter-sound correspondences is essential for developing fluent decoding skills (Adams, 1990), but decoding ability alone will not lead to successful reading (Ehri, 1997). Ehri (1989) pointed out that decoding skills facilitate the learning of letter patterns within word spellings (both phonemically regular and irregular) that occur across different words and the phonemes these patterns represent. Knowledge of letter patterns in spellings has been shown to correlate strongly with reading fluency, likely because knowledge of letter patterns facilitates sight-word knowledge (Juel, 1988). Another important aspect of alphabetic understanding is the ability to blend phonemes represented in printed words. Frequently, this ability is measured with tasks of non-word reading. Speece, et al. (2003) found that a measure of non-word reading administered in kindergarten correlated to both a measure of oral reading fluency ($r = .69$) and the Woodcock Johnson-Revised Letter-Word ID subtest ($r = .76$) (Woodcock & Johnson, 1990).

**Fluency.** The term “fluency” (or automaticity) is defined in a variety of ways when pertaining to reading. In general, fluency refers to the demonstration of both speed and accuracy in performing a task. According to the National Reading Panel (2000) fluent reading requires,
“the ability to read text quickly, accurately, and with proper expression” (p. 3-5). Ehri (1989) characterizes the ability to read words fluently and by sight as the key to skilled reading.

LaBerge and Samuels (1974) explained the connection between fluency and effective reading with their theory of automatic information processing. When words are recognized with little effort, cognitive attention is reserved for comprehension of meaning. Likewise, fluent word recognition rests on the development of fluency in reading sub-skills (e.g., identifying letter-sound correspondences). Speece et al. (2003), in a study of the influence of letter fluency tasks on later reading, distinguished fluency tasks from Rapid Automatized Naming (RAN) tasks (Denkla & Rudel, 1976; Wolf & Bowers, 1999) by defining fluency in terms of “the speed and accuracy with which multiple exemplars can be produced orally” (p. 223). The key difference is that fluency tasks, when used as indicators of reading skill, require fluent recognition of multiple stimuli (e.g., all letters in the alphabet) whereas RAN tasks rely on a handful of known stimuli (e.g., a limited number of colors or objects) (Speece et al., 2003).

Fluency is an important component of reading, “that pervades all levels of processing involved in reading and…fluency on early foundational skills can be used to predict proficiency in subsequent [more complex] skills” (Good, Simmons, & Kame’enui, 2001, p. 264). Supporting this notion, many studies have demonstrated that fluent word recognition and oral reading skills are strongly associated with reading comprehension (e.g., Deno, Mirkin, & Chiang, 1982; Good & Jefferson, 1998; National Reading Panel, 2000; Shinn & Good, 1992; Stanovich, 1985). Studies have also demonstrated the effective use of fluency-based measures of reading sub-skills to predict a student’s risk for future reading failure (e.g., Fuchs, Fuchs, & Compton, 2004; Good, & Harn, 2001; Good et al., 2000; Kaminski & Good, 1996; Speece, et al. 2003).
sum, the general inference can be made that fluency with particular reading skills can represent the mastery and integration of all previous and less complex skills.

**Conventions of Print.** The construct of “conventions of print” (Reid, et al., 2001) or “concepts of print” (National Reading Council, 1998) refers to an understanding of how print is used, rather than specific knowledge or skill with using print (e.g., knowledge of letter-sounds, writing). Understanding conventions of print requires a child to have some rudimentary understanding that print is used in particular rule-governed ways (e.g., books have a proper orientation, print is read left to right, pages turned left to right), and that print is used to communicate information (e.g., on logos, street signs, food boxes; in books) (Reid et al., 2001). Scarborough (1998), reviewed seven predictive studies and found that concepts of print had a moderate mean correlation ($r = .49$) with later reading difficulty. Stuart (1995) used a measure of concepts of print, combined with a measure of letter-naming and found strong correlations to reading in the early elementary grades. None of the DIBS investigated in this study measures conventions of print as defined here. However, Good and Kaminski (1996) suggest that DIBELS Letter-Naming Fluency is a measure of “awareness of print”. Although the strong relationship between letter-naming tasks and later reading performance may be due to the similarity of letter-naming to letter-sound identification, it is possible that letter-naming tasks are also related to a child’s knowledge of conventions of print. Conventions of print and letter-naming fluency in kindergarten may capture unique aspects of a larger construct that relates to a child’s overall literacy exposure and experience with print. In this study, students’ knowledge of conventions of print was measured by the TERA-3 Conventions subtest (described in Methods section), and is used as an outcome measure in the spring of kindergarten.
**Dynamic Indicators of Basic Early Literacy Skills.** DIBELS is a published set of DIBS designed to measure early literacy skills that are predictive of later reading competence. DIBELS research began with the use of CBM-ORF, which later became DIBELS-ORF to predict reading performance. Gradually other subtests were incorporated that encompassed the early literacy skill domains thought to be important to successful reading acquisition. Good, Simmons, and Kame’enui (2001) indicated that "selected foundational skills include (a) phonological awareness or the ability to hear and manipulate the sound structure of language, (b) alphabetic understanding or the mapping of print to speech and the phonological recoding of letter strings into corresponding sounds and blending stored sounds into words, and (c) accuracy and fluency with connected text or the facile and seemingly effortless recognition of words in connected text" (p. 261). Within DIBELS, these skills are presumed to lie on a continuum, progressing in complexity until they culminate in the ultimate goal of actual reading (i.e. constructing meaning from a variety of texts) usually by the end of the third grade (Good et al., 2001). These skills do not represent all important early literacy skills, but they are thought to be valid indicator skills that predict performance on subsequent and more complex reading skills. DIBELS measures of phonemic awareness include Initial Sound Fluency (D-ISF) and Phonemic Segmentation Fluency (D-PSF). The DIBELS measure of letter-knowledge is Letter Naming Fluency (D-LNF). DIBELS measures of alphabetic understanding (i.e., letter-sound correspondence, phoneme blending) is Nonsense Word Fluency (D-NWF). The final DIBELS measure of fluent word recognition within connected text is Oral Reading Fluency (D-ORF).

For screening and benchmark progress monitoring, a set of measures is administered to each grade level—pre-school through third grade—at the beginning, middle, and end of the academic year. At each of these benchmark administrations, new subtests may be added and
others dropped, depending on the particular predictive power of a subtest at a particular scheduled administration (Good, Gruba, & Kaminski, 2001). Each subtest also changes in its predictive value—a subtest that indicates a student is at risk in kindergarten may not accurately predict risk if it were administered in the first grade (Good et al., 2001). For example, measures of phonemic awareness skill are hypothesized to be most predictive when students are in kindergarten, before formal instruction in higher-level reading skills begins (Good et al., 2001; O’Connor, 2000). DIBELS uses Initial Sound Fluency (D-ISF) and Phoneme Segmentation Fluency (D-PSF) subtests to assess phonemic awareness during kindergarten. In the beginning of first grade, performance expectations expand to include alphabetic understanding—the understanding of fundamental written language principles. Aspects of alphabetic understanding are assessed with Letter Naming Fluency (D-LNF) and Nonsense Word Fluency (D-NWF). In the middle of first grade, after formal reading instruction has begun, Oral Reading Fluency (D-ORF) is introduced to assess a student’s emerging ability to read words in connected text. By the second grade, fluency with connected text becomes the strongest indicator of general reading competence. Because fluent text reading is a more complex skill that requires the integration of earlier component literacy skills, it is assumed that measurements of D-ORF represent the range of prerequisite skills targeted by earlier subtests (Kaminski & Good, 1996; Good et al., 2001). D-ORF is thus the primary indicator of risk status and reading competence used after the first grade (Good et al., 2001).

As described earlier, DIBELS measures, like all DIBS, differ from CBM in important ways. One key difference is that DIBELS measures target skills that are not usually taught in the classroom. In kindergarten, for example, explicit instruction in rapid letter-naming and nonsense-word reading are not included in most standard curricula. Despite this, Letter-Naming Fluency
and Nonsense Word Fluency are strong indicators of future reading performance. This logic can cause confusion for practitioners who fail to understand the connection between the skills measured and the implications for instruction (Good et al., 2001). Furthermore, the skills measured by DIBELS are often confused as goals for instruction. Creating instructional lessons based on improving the skills measured by DIBELS is not necessarily productive: Teaching fluency with letter names will not make students better readers, but students’ skill with letter naming is a powerful indicator of risk-status nonetheless (Speece et al., 2003). To clarify both of these sources of confusion, users of DIBELS and other DIBS must understand the measurement logic of using component skills to indicate broader competence, and they must understand how the interpretation of DIBS scores can lead to improved instructional outcomes.

**DIBELS within an Outcomes-Driven Model of Educational Decision-Making**

DIBELS measures are used within a prevention-oriented, Outcomes-Driven Model proposed by Good et al. (2001), which is similar to the problem-solving model used for making decisions with CBM (Deno, 1989; Shinn, 1995). The goal of CBM, used within the problem-solving model is the identification and remediation of existing academic problems. In contrast, the Outcomes-Driven Model was developed as a framework for assessment and intervention decision-making, focused on the prevention of early reading failure and on the outcome of adequate reading achievement (Good et al., 2003). Five core decisions lead step-by-step to this goal: (a) identifying the need for support, (b) validating the need for support, (c) planning support, (d) evaluating and modifying support, and (e) reviewing outcomes.

**Identifying the need for support.** This first step is intended to identify students in need of support through school-wide benchmark assessment. Three times per year (fall, winter, and spring), for the purposes of screening and benchmark progress monitoring, all students are
administered the scheduled DIBELS subtests. Students who are identified as at-risk are provided with intervention designed to help them achieve subsequent benchmark goals (Good et al., in press).

*Validating the need for support.* This step involves eliminating other factors from consideration that might contribute to low performance. This involves repeated administrations of the subtests under different conditions. If subsequent scores are similar to those from initial administrations, the need for support is demonstrated.

*Planning instructional support.* When the need for support is validated, students are immediately provided with appropriate instruction that supports learning necessary early reading skills. Although the type of instruction is not specified within the Outcomes-Driven Model, numerous research-based strategies are available that are effective and appropriate for use in this step.

*Evaluating outcomes.* Evaluating Outcomes involves determining whether research-based instructional interventions are effective at improving outcomes. Students who are receiving more strategic and intensive instructional support will require assessment that is more frequent. DIBELS is suitable for this purpose because the subtests have numerous alternate forms and are sensitive to changes over short periods of time. To evaluate the effectiveness of supports for a student weekly assessments are administered. If progress is being made towards the next benchmark goal, the support is continued. If assessment indicates that progress is not being made, the instructional support may be modified or intensified (Kame’enui & Simmons, 1998).

*Reviewing outcomes.* This step exemplifies the potential to use DIBELS assessment data for multiple decisions. Data from benchmark assessments are used to review several outcomes.
For individual students, a determination is made as to whether the student is making appropriate progress towards benchmark goals. If not, the model directs the process back to the planning instruction step, and a more intensive review of instructional options is conducted. To review outcomes for groups of students, schools examine benchmark data for all students assessed. Comparisons in the level and rate of progress can be made to local and national norm groups, other schools within the district, or past performance at the same school. Likewise, if progress towards school-wide goals is not on-track, the planning support step is repeated with the identification of school-wide instructional supports the goal.

Research Evaluating the Use of DIBELS in Kindergarten

The use of DIBS in early literacy, like those within the DIBELS package, represents a set of relatively new measurement practices. Unlike that of CBM, the research evaluating the technical adequacy of early literacy measures is limited. Only a handful of studies have been conducted that directly examine the reliability and validity of DIBELS and provide initial support for their use in Outcomes-Driven Models. Table 1 summarizes estimations of the reliability of DIBELS measures. Table 2 summarizes evidence supporting various estimations of validity for DIBELS.

In one of the first validation studies of DIBELS, Kaminski and Good (1996) examined the concurrent and predictive validity of DIBELS Letter Naming Fluency (D-LNF) and Phoneme Segmentation Fluency (D-PSF) for 18 kindergartners. Significant moderate correlations were found between kindergarten scores and the McCarthy Scales of Children’s Abilities (McCarthy, 1972) the Metropolitan Readiness Test (Nurss & McGauvran, 1986), and the Stanford Diagnostic Reading Test (Karlsen & Gardner, 1985). However, D-LNF scores for first-grade significantly correlated with only the Stanford Diagnostic Reading Test ($r = .72$) while D-PSF
scores for first-grade did not correlate significantly with any criterion measures. This study provided initial evidence that D-LNF was a valid predictor of reading.

Hintze, Ryan, and Stoner (2003) examined the concurrent validity of the DIBELS measures and the Comprehensive Test of Phonological Processing (CTOPP) (Wagner, Torgesen, & Rashotte, 1999). Strong correlations were found between kindergarten D-PSF, D-LNF scores, and the subtests of the CTOPP measuring similar skills. Hintze et al. concluded that the coefficients in their study suggested that D-ISF and D-PSF tasks are more closely associated with those subtests and composite scores of the CTOPP that measure phonological awareness and less associated with tasks involving rapid naming. They also concluded that D-LNF was associated with CTOPP subtests that involved both phonological awareness and rapid naming. Again, D-LNF was shown to be related to multiple skill domains measured by criterion subtests.

Elliot et al. (2001) examined the criterion-related validity of D-LNF and D-PSF subtests and two modified measures of initial phoneme ability and phoneme segmentation ability. The authors eliminated the fluency component of the D-ISF and D-PSF subtests and modified items to increase in difficulty. The number of items answered correctly was calculated as the students’ score. The authors suggested that these modifications resulted in measures of ability rather than fluency, and compared the correlation coefficients of ability and fluency measures to criterion measures of reading. Scores from 75 kindergarteners’ Letter Naming Fluency (D-LNF) and phoneme-segmentation ability (similar to D-PSF) meaningfully predicted scores on the Woodcock-Johnson Psycho-Educational Battery-Revised, Broad Reading Cluster (Woodcock & Johnson, 1990) and the Test of Phonological Awareness (Torgesen & Bryant, 1994). The phoneme-segmentation ability measure correlated with the WJ-R Broad Reading Cluster \((r = .44)\) and the Letter-Word Identification subtest \((r = .45)\), but these coefficients were weaker than
were found for D-PSF and reading measures in previous studies (Good, et al., 2000; 2001). The initial phoneme ability measure correlated with the WJ-R Broad Reading Cluster ($r = .42$) and the Letter-Word Identification subtest ($r = .47$), which are slightly stronger than correlations between D-ISF and criterion measures previously found by Good, et al. (2001). Though inferences are difficult to make from the comparison of only two studies, the slightly superior performance of an initial phoneme ability measure indicates the need for further investigation of the technical characteristics of D-ISF, especially with other outcome measures of reading. An interesting finding is that D-LNF and an experimental measure of sound-naming fluency (similar to A-LSF) correlated with the WJ-R Broad Reading Cluster ($r = .63$, $r = .58$ respectively) and with the WJ-R Letter-Word Identification subtest ($r = .71$, $r = .62$) more strongly than either of the experimental ability measures. These strong correlations between D-LNF and reading outcomes in this study are consistent with previous studies that have demonstrated the relationship between letter-naming fluency and reading measures. DIBELS does not include a measure of letter-sound fluency, and the correlations of the sound-naming ability measure with criterion measures suggests that letter-sound fluency may be an important skill to measure in the early school years. The present study extends the results of Elliot et al. (2001) by including A-LSF along with the DIBELS measures administered throughout the kindergarten year. Of particular interest was whether A-LSF would tap critical skills that are not measured by the standard DIBELS battery, and whether A-LSF would explain part of the strong relationship between D-LNF and reading outcomes demonstrated throughout the literature.

VanDerHeyden et al. (2001) examined the reliability and validity of a set of fluency-based CBM readiness probes for kindergarten that would identify students in need of intervention in reading, writing and math. These probes have technical characteristics similar to
those of the other DIBS in early literacy as described in this study, and could be conceptualized as DIBS. The authors administered the experimental measures to 107 students in kindergarten. The experimental measures for math included counting skills and number identification. Reading was assessed by measuring students’ fluency with identifying initial word sounds and corresponding written letters. Writing was measured with letter writing fluency. Alternate forms reliability coefficients ranged from .58 to .84 for all measures, and inter-rater reliability ranged from 96% to 100%. As criterion measures, 9 subtests of the Readiness Assessment section of the Comprehensive Inventory of Basic Skills (Brigance, 1999) were used as well as D-ISF. Results showed significant correlations between the experimental letter-writing measures and the subtests of the CIBS-R that required letter reading. Interestingly, the experimental math probes also correlated significantly with CIBS-R letter reading tasks as well. All of the experimental measures (reading, writing, and math) correlated with the CIBS-R subtest that required reading upper-case letters. This subtest is similar to D-LNF, that is conceptualized as a measure of general risk rather than a measurement of a critical skill. These results add to the empirical evidence supporting the use of letter-naming tasks like D-LNF as predictors of reading, and suggest that letter-naming could be linked to a larger construct that connects reading, writing, and math together.

Burke, Hagan-Burke, and Ditkowsky (in review) investigated the predictive validity of D-ISF and D-LNF with later reading outcomes, when administered to a sample of 129 children at the beginning of kindergarten. Strong correlations were found between D-LNF and D-ORF at the middle ($r = .65$) and the end ($r = .62$) of first grade. Correlations between D-ISF and D-ORF were considerably weaker at both points (middle of first, $r = .29$; end of first, $r = .23$).
Consistent with previous studies, these results demonstrate that letter-naming tasks administered in kindergarten are more predictive of later reading than measures of phonological awareness.

That a strong relationship exists between D-LNF and later reading competence seems clear. What remains unclear, however, is exactly what construct D-LNF measures considering the strong correlations to other skills like number-identification (VanDerHeyden et al., 2001; Vellutino & Scanlon, 1996). The possibility exists that D-LNF measures an unknown construct that governs the risk of low achievement in general. If this is indeed true, and broader construct is being tapped by D-LNF, other potentially irrelevant aspects of that construct may influence D-LNF scores. Although D-LNF identifies students on the path to reading failure, it may also identify students as at-risk who will go on to become successful readers without additional intervention. Clarifying whether D-LNF is over-predictive of risk-status will be an important goal for future research.

The problem of over-identification has been noted in other studies when phonological awareness skills were used to identify students at-risk. Measuring phonological awareness skills in kindergarten has repeatedly been shown to over-identify students with poor phonological skills who later go on to become successful readers without additional intervention (Felton, 1992; O’Connor & Jenkins, 1999; Scarborough, 1998; Wagner, 1997; Uhry, 1992). For example, most children do not enter kindergarten with an awareness of phonemes and the ability to segment words into phonemes—phonemic awareness almost certainly requires instruction to develop (Adams, 1990; Torgesen, 2000). As the results found by Hintze et al. (2003) illustrate, many students identified for intervention with phonemic segmentation measures will respond favorably to instruction and acquire the skill. Although over-predicting students for intervention ensures that most students who need intervention will receive it, providing additional
intervention to many students who do not require it is a clear waste of resources, and depending on the quality of the intervention, may be harmful (O’Connor & Jenkins, 1999). Delaying the time of screening for reading difficulties (e.g., until first grade) would result in more accurate identification, but this risks delaying intervention for students who would benefit from it in kindergarten (O’Connor & Jenkins, 1999). Early and accurate identification of students at-risk for reading difficulties is clearly a priority.

To maximize the accuracy of DIBELS for the purposes of classification, Good et al. (2002) established cut scores for each measure that correspond to categories of risk. An effectively chosen cut score can be used to classify students into categories (e.g., at-risk for reading failure, not-at-risk for reading failure) when compared to classifications made using a criterion measure that is presumed to be valid (e.g., actual reading achievement one year later). From this comparison, the number of true positives and negatives (i.e., predictor and criterion agreement on the presence or absence of a problem) and false positives and negatives (predictor and criterion disagreement as to the presence or absence of a problem) is determined. These data are used to estimate sensitivity, specificity and the positive and negative predictive power for a measure. Sensitivity refers to the likelihood that a predictor measure will identify the same students as at-risk as are identified by the criterion measure (i.e., true positives). Specificity refers to the likelihood that the predictor measure will identify students that have been identified by the criterion measure as not-at-risk (i.e., true negatives). Positive predictive power refers to the likelihood that students identified by a predictor as being at-risk are supported by the criterion measure. Likewise, negative predictive power refers to the likelihood that students identified by the predictor as not-at-risk will be supported by the criterion measure (Swets, Dawes, & Monohan, 2000). Sensitivity and specificity are inversely related; an increase in one
results in a decrease in the other. Predicting these values accurately has both educational and cost-benefit implications. When this logic is applied to measures of early literacy used for the purposes of identification for additional services, the goal is to balance the number of false positives and false negatives so that resources spent are not on students who are not in need of intervention, and that as many students as possible who are genuinely in need of intervention are identified.

Good et al. (2002) based the establishment of cut scores for subtests on the conceptualization of early reading acquisition as a continuum of skills that progress step-wise, one to the other, and increase in complexity (i.e., phonological awareness, alphabetic understanding, and automaticity with reading connected text) (Kaminski & Good, 1998). Thus, a subtest that measures a particular early literacy skill should predict performance on a later skill. Risk categories based on a range of critical scores for each measure. The scores within each category indicate the likelihood of a students’ performance on the same measure at the next benchmark. For example, a student who scores less than 4 on D-ISF in the fall of kindergarten is placed in the “at-risk” category, indicating that without intervention, this student is at-risk for failing to reach the score of 25 on the winter D-ISF—the minimum score designating the “low-risk” category. Table 3 summarizes the cut scores established with this procedure for students in kindergarten and the risk status associated with each range scores. Good et al. (2002) showed that by moving cuts cores for DIBELS subtests up or down, the accuracy of the predictions of future performance on other subtests could be maximized.

In the study by Hintze et al. (2003) discussed previously, the authors also examined the diagnostic accuracy of the established cuts scores for DIBELS kindergarten subtests, using the CTOPP as the criterion measure instead of DIBELS subtests at later benchmarks (at-risk
standard scores on CTOPP subtests were those below 85 points, or one standard deviation below the mean). The authors found that predicting performance on the CTOPP based on established cut scores for D-ISF and D-PSF resulted in dramatic over-identification of students as at-risk (when risk is defined by performance on the CTOPP). The authors note that the percentage of students classified into risk categories accurately with DIBELS cut scores was not significantly better than chance.

**Rationale and Research Questions**

Students at-risk for reading failure who are not identified early may be identified later when reading problems are evident, but by this time, options for effective intervention will be limited. Kindergarten represents a critical window of opportunity when the effects of instructional interventions result are maximized for struggling readers (Adams, 1990; National Reading Panel, 2000; Snow, Burns, & Griffin, 1998). Identifying students in need of support, the first step of the Outcomes-Driven Model (Good et al., 2001), requires measurement tools that are reliable and valid. To this end, the present study was intended to extend previous research by examining the reliability and validity of a set of DIBS in early literacy, specifically DIBELS and A-LSF, used to identify students at-risk for reading failure in kindergarten.

A consistent finding reported across the research literature on DIBELS is that D-LNF is the strongest predictor of reading when administered at the beginning of kindergarten (Good et al., 2000; Hintze et al., 2003, Van Der Heyden et al., 2001; Elliot, et al., 2001; Speece, et al., 2003). D-NWF administered at the end of kindergarten demonstrates somewhat stronger predictive validity than D-LNF (Speece et al., 2003) Generally weaker correlations are found between D-ISF and D-PSF with reading and the other measures involving print (Good et al., 2001; Hagan-Burke et al., in press; VanDerHeyden et al., 2001). This is somewhat surprising
considering the theoretical and empirical support for the predictive power of measures of phonological awareness (Adams, 1990; National Reading Panel 2000; National Research Council, 1998). The theoretical underpinnings of DIBELS also place central importance on the development of phonological awareness skills for reading acquisition, the differential support for print-related measures brings this perspective into question. It is possible that measures of phonological awareness and print-related skills are more or less predictive based on when they are administered.

The present study examined whether changes were evident in the predictive validity coefficients of DIBELS and A-LSF when administered at different times during the kindergarten year. The strong relationship of D-LNF to reading is also not fully understood. Within DIBELS, D-LNF is described as a “measure of risk,” for reading failure. However, this label does little to explain why D-LNF is so predictive of reading in light of the fact that letter-naming skills are not necessary for reading acquisition. Because of the potential connection of D-LNF to letter-sound skills (Ehri, 1989; Shinn, 1995), AimsWeb Letter-Sound Fluency (A-LSF) (Shinn & Shinn, 2002) was included as a kindergarten measure in the present study to explore whether some of the correlation expected between D-LNF and reading would be explained by A-LSF. In addition, the study evaluated whether A-LSF alone was an adequate predictor of reading that would contribute meaningfully to the DIBELS measures indicated for kindergarten. Two criterion measures of reading were chosen for administration in the spring—the TERA-3 and D-ORF. The skills measured by the TERA-3 and D-ORF represented a relatively broad range of meaningful reading outcomes for students at the end of kindergarten that were thought to be suitable as dependent measures. Correlation and multiple regression were used to examine the following questions regarding reliability and validity:
Reliability

Question 1. Will scores from identical forms of DIBS (DIBELS subtests and A-LSF) administered two weeks apart correlate strongly \((r > .80)\), demonstrating that the skills measured are stable? The DIBS investigated in this study are designed to be sensitive to changes in performance over a relatively short period. If sufficient time elapses between repeated measurements, changes in scores are expected. The difference between DIBS scores administered at the beginning of kindergarten and at the middle will almost certainly be significant. The distribution of scores and the overall variance for a group might also change. Administering identical forms of DIBS after an interim of two weeks was not thought to be enough of an interim to produce considerably different scores. The expectation is that scores from the two administrations will correlate very strongly; correlation coefficients of the DIBS are expected to be in the range of .80 to .90.

Predictive Criterion-Related Validity

Question 2. What are the correlations between DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF), at the middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF), and at the end (D-LNF, D-PSF, D-NWF, and A-LSF) of kindergarten? This question is intended to investigate the power of DIBS administered at one point in time predicting scores on DIBS at a later point in time. Good and Kaminski (2002) recommend the use of specific DIBELS measures at each kindergarten benchmark administration, based on a “stepping-stone” model of reading acquisition. This model places initial importance on the development of phonological awareness skills that lead to the development of alphabetic understanding. Based on this model, the measures tapping phonological awareness (D-ISF and D-PSF) should correlate at least moderately with one another and with the measures involving letter-sound correspondence (D-
NWF, A-LSF) at subsequent administrations. However, considering the evidence from the literature that has demonstrated mostly weak correlations between D-ISF and later measures of reading (Burke et al., in review; Good et al., 2001), it was not clear whether D-ISF would perform as the strongly as the DIBELS model implies. Within the same model, D-LNF is considered a measure of risk, and as such, should correlate at least moderately with later measures involving skills with print (D-NWF, A-LSF). Although considering the strong correlations between D-LNF and outcome measures of reading demonstrated in the literature, D-LNF was hypothesized to produce moderate to strong correlations with other measures of print. The correlations between A-LSF and other measures were difficult to predict. Since most students at the beginning of kindergarten do not have knowledge of letter-sound correspondences, the scores from the fall administration were expected to be very low, and weakly correlated with later measures. However, A-LSF from winter was expected to correlate to D-NWF since both are measures of letter-sound correspondence and both develop as a result of instruction.

Question 3. What are the correlations between DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF) and middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF) of kindergarten, and criterion measures of reading ability administered in the spring (TERA-3, D-ORF)? Similar to Question 2, developing hypotheses regarding the strength of the relationships between the DIBS in early literacy and the outcome measures depends a great deal on theoretical orientation. Following the stepping stone model of reading acquisition suggested by Good et al., (2001), measures of phonological awareness from the beginning (D-ISF) should produce the strongest correlations with criterion measures of the DIBS given in the fall. In the winter, (D-PSF) represents a more sophisticated level of phonemic awareness and should correlate
moderately to criterion measures. Likewise in the winter, after skills with print have begun to emerge, D-NWF and A-LSF should demonstrate moderate correlations with criterion measures. D-LNF, as a measure of risk, should correlate moderately with criterion measures in both the fall and the winter.

The hypothesized answers to this question drew from the DIBELS model of reading described above, but were also influenced by evidence from research, indicating that measures involving print administered in the fall and winter are superior predictors of later reading than measures of phonological awareness (e.g., Elliot et al., 2001; Speece et al., 2003). Thus, D-ISF was predicted to correlate weakly to moderately with criterion measures of reading at the end of kindergarten. Moderate correlations were predicted for winter D-PSF and criterion measures. D-LNF was predicted to have the strongest correlations with criterion measures of any of the fall measures. Winter D-LNF was predicted to correlate with criterion measures less strongly because of ceiling effects for this measure: By the middle of kindergarten, more children have learned letter names, and D-LNF may not discriminate between children as well. In the fall, A-LSF was predicted to correlate weakly to criterion measures because of anticipated floor effects. Winter A-LSF, on the other hand, was predicted to correlate with criterion measures more strongly than in fall since the winter administration was thought to have reflected the effects of instruction. Likewise, D-NWF in the winter should also correlate moderately to strongly with criterion measures.

**Question 4.** Will particular DIBELS measures, of those administered at the beginning and at the middle of kindergarten, explain significant unique variance in criterion measures of reading administered at the end of kindergarten (TERA-3 and D-ORF)? This question focuses specifically on the DIBELS measures administered in the fall and winter of kindergarten.
Although each of the measures is predicted to explain some amount of variance in the criterion measures, predicting whether a particular measure will uniquely account for a significant amount of variance requires considering both evidence from research as well as theory. Disagreement exists regarding the role phonological skills play in the process of early reading acquisition. From one perspective, phonological awareness is hypothesized to be a prerequisite to acquiring alphabetic understanding, also a necessary prerequisite skill. Based on this logic, D-ISF administered at the beginning of kindergarten should explain a meaningful amount of the variance in the criterion measures or measures that serve as mediators between D-ISF and the criterion measures. D-PSF is a measure of phonemic awareness and in the winter should explain a unique proportion of variance in criterion measures. An alternative theoretical perspective suggests that phonological awareness are not critically necessary for reading acquisition, and rather develop as a result of reading instruction and acquiring skills with print (Ehri, 1989). The implication is that measures of phonological awareness skill may be correlated with criterion measures of reading, but will not likely explain unique variance. Measures of skills with print like D-NWF would likely explain unique variance in criterion measures. This theoretical perspective is somewhat more consistent with the research, that indicates measures involving print are more strongly predictive of later reading outcomes than measures of phonological awareness (Hammill, 2004; Scarborough, 1998; Schatschnieder et. al, 2004; Speece et al, 2003). In this way, measures involving print would likely explain more unique variance in criterion measures as well.

For the present study, D-ISF was hypothesized to account for a small amount of the variance in criterion measures. Based on the evidence of previous studies, fall D-LNF was predicted to account for significant variance in the criterion measures after controlling for D-ISF,
This would support the use of D-LNF as a predictor of risk in the fall of kindergarten. In the winter, D-LNF and D-NWF were predicted to account for unique variance in criterion measures. It is possible that D-NWF will account for a greater amount when effects of other variables, specifically D-LNF, are controlled. If found, this would support the perspective that print-oriented skills are critically important for learning to read in kindergarten and would support their use as predictors of reading and would indicate the need for closer examination the relationship

*Question 5. Will A-LSF, administered at the beginning and at the middle of kindergarten explain significant unique variance to criterion measures of reading (TERA-3 and ORF) after controlling for the effects of DIBELS Subtests?* DIBELS does not include a specific measure of letter-sound fluency. Although D-NWF requires understanding letter-sound correspondence, the administration directions (described in the next chapter) indicate that the sounds should be blended together if possible. Thus, for some students, D-NWF represents both letter-sound knowledge and sound blending ability. A-LSF requires students to demonstrate the phoneme from written letters on a page, but in isolation. Logically, children will learn letter-sounds in isolation before they learn to blend them into strings of phonemes—both important skills (Ehri, 1985, 1989). If letter-sound knowledge indeed develops in this sequence, A-LSF (specifically in the winter) could account for a unique amount of variance in criterion measures that D-NWF misses because of floor effects. By spring, when many students have developed skills with letter-sounds and sound blending, D-NWF may predict more variance than A-LSF. Also of interest was whether the amount of variance explained in fall and winter by D-LNF will change when the effects of A-LSF are controlled. If this effect is found, some light may be shed on the nature of the strong relationship between D-LNF and later reading outcomes.
Concurrent Criterion-Related Validity

Question 6. What are the correlations between DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) and criterion measures of reading (TERA-3 and ORF)? All of the DIBS administered in the spring are hypothesized to correlate at least moderately with both the TERA-3 and D-ORF. The measures involving print are predicted to have stronger correlations than D-PSF since the outcome measures (TEAR-3 and D-ORF) both involve skills with print similar to those measured by the other DIBS in spring. D-NWF is predicted to correlate strongly to criterion measures, while A-LSF and D-LNF are predicted to correlate somewhat less strongly than D-NWF.

Question 7. Will certain DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) explain unique amounts of variance of criterion measures of reading (TERA-3 and ORF)? In previous studies, D-LNF, when administered in the fall and winter of kindergarten, demonstrates consistently high correlations with criterion measures of reading; However, in the spring of kindergarten, the strength of the correlations between D-LNF and reading measures diminishes somewhat, and D-NWF demonstrates stronger correlations with criterion measures (Elliot et al., 2001; Hintze et al., 2003; Speece et al., 2003). Since most children have learned letter-names by the end of the kindergarten year, D-LNF can produce a ceiling effect in the spring (Schatzschneider et al., 2004). Therefore, D-LNF is not hypothesized to account for a significantly greater proportion of variance in criterion measures than other DIBS in spring. D-NWF is the most complex of the DIBS administered in the spring, and it is likely that children who are competent with D-NWF have integrated the skills measured by A-LSF, D-PSF, and D-LNF. Thus, D-NWF was hypothesized to account for a large amount of the variance in the criterion measures.
CHAPTER 3

Method

Participants

For the present study, sample data for 177 kindergarten students were complied from an extant database of performance scores on measures of early literacy skills. The students represented in the data were enrolled at a rural primary school in Georgia that administered DIBS in early literacy and norm-referenced tests of reading achievement three times yearly, (i.e., fall, winter, and spring). Both types of measures were used for screening and progress monitoring purposes. Parents provided consent for the school to assess students, and understood that the data were used for the purposes of instructional planning and identification of students in need of additional academic support. For the purposes of this study, the school provided consent for the analyses of the data.

Of 225 students enrolled at the end of the year, 180 had been assessed on all of the measures analyzed in this study. Three of the remaining 180 students scored extremely high on the study measures compared to the average scores (i.e., above the 95th percentile on more than half of the measures). The scores of these three students were removed in order to normalize the distribution of scores. Table 4 summarizes relevant participant characteristics. Eighty-nine of the students were male and 88 were female. The mean age of the students in the sample was five years eight months. Approximately 60% (n = 107) were Caucasian, 34% were African-American (n = 60), 1% were Hispanic (n = 2), 4% were Multi-Ethnic (n = 7), and one student was Asian. Subsidized school lunch was used as an indicator of socio-economic status: Thirty-seven percent
(n = 66) of students received free school lunch, approximately 3% (n = 5) received partial financial assistance, and 60% (n = 106) received no financial assistance. Twenty-seven students (15%) received some type of special instructional services (e.g., speech and language, early intervention, learning disabilities, emotional behavior disorders). No students were eliminated from the sample because of low scores or a documented disability. School administrators reported that for the past three years, the school has met the statewide goal of at least 95% students passing yearly competency tests mandated by the State of Georgia.

Description of Measures

A set of DIBS in early literacy were administered individually in the fall, winter, and spring. Table 5 summarizes the schedule of administration of each measure. Descriptions of the DIBS in early literacy (DIBELS kindergarten measures and AimsWeb Letter-Sound Fluency) are provided along with reliability and validity evidence reported in previous research (see Tables 1 and 2 for a summary of coefficients). DIBELS-Oral Reading Fluency (Kaminski & Good, 2002) and Test of Early Reading Ability-Third Edition (Reid, Hresko, & Hammill, 2001) were administered in the spring as criterion measures. Raw scores from all measures were used in the correlation and regression analyses.

Predictor Measures

DIBELS Initial Sound Fluency (D-ISF) (See Appendix A). D-ISF is a measure of phonological awareness, specifically knowledge and awareness of the initial sounds in words presented orally (Kaminski & Good, 1998). Students are presented with four pictures and the examiner identifies each picture by pointing and naming them. The student is then asked to identify by pointing, which picture begins with a particular sound. For example, if the examiner says, “This is bee, lamp, cake, and sink. Which picture begins with /b/?” and the child points to
the picture with the correct initial sound. Twenty opportunities are given for the student correctly to identify an initial sound. The examiner measures the time between the presentation of the question (“Which picture begins with /b/?”) and the student’s correct identification of the initial sound. The total time taken as well as the number of sounds out of twenty that were correctly identified is calculated and these are converted into the number of correctly identified sounds per minute. This rate is the student’s score for D-ISF. Equivalent forms reliability estimates for D-ISF range from .82 (Elliot et al., 2001) to .91 (Good and Kaminski, 1996). Concurrent validity estimates for D-ISF with the WJ-R Reading Cluster Score is .42, and with the Test of Phonological Awareness is .62 (Elliot et al., 2001).

*DIBELS Phoneme Segmentation Fluency (D-PSF)* (See Appendix B). D-PSF measures a more complex aspect of phonological awareness, phonemic awareness (Good and Kaminski, 1996). Examiners read aloud words of three to four phonemes and ask the student verbally to produce each phoneme within the word. Each phoneme spoken correctly is scored as a point. For example, if the examiner says “dog,” the student says “/d/ /o/ /g/” to receive three points for this word. The total number of correctly spoken phonemes in one minute is calculated as the final score. Test-retest reliability for D-PSF is .85 (Elliot et al., 2001). Equivalent forms reliability ranged from .79 (Good & Kaminski, 1996) to .88 (Good et al., 2000). Kindergarten scores on D-PSF were correlated with scores from 1st grade Woodcock-Johnson Revised Reading readiness cluster resulting in a validity coefficient of .54 (Good et al., 2001). Kindergarten D-PSF scores were compared to scores on the Metropolitan Readiness resulting in a validity coefficient of .68 (Good et al., 2000).

*DIBELS Letter Naming Fluency (D-LNF)* (See Appendix C). D-LNF, which is used from beginning kindergarten through the beginning of first grade, is not a measure of phonological
awareness, but of letter knowledge. Although letter knowledge is not a necessary skill for learning to read, studies have shown that it is a powerful indicator of risk (Kaminski & Good, 1996). The reason for this is not fully understood, but one explanation is that letter knowledge is closely related to letter-sound knowledge, an important component of alphabetic understanding. Alphabetic understanding involves children learning that printed letters correspond to sounds, and this is a critical step in learning to read. (For this subtest, the examiner presents the student with a list of upper and lower-case letters. The student is asked to identify orally the names of as many letters as they can and are stopped after one minute. The score is the number of correctly named letters in one minute. Reliability estimates for D-LNF range from .88 (Good & Kaminski, 1996) to .93 (Good et al., 2000). Criterion validity of first grade D-ISF scores with DIBELS Oral Reading Fluency is .65 (Good & Kaminski, 1996) and with the Test of Phonological Awareness is .50 (Elliot et al., 2001).

**DIBELS Nonsense Word Fluency (D-NWF)** (See Appendix D). D-NWF is a measure of the alphabetic principle, specifically a student’s ability to identify letter-sound correspondences and to blend letter sounds into words (Kaminski & Good, 1996). The student is presented with a list of vowel-consonant and consonant-vowel-consonant nonsense words (e.g., wuj, ig, tiv), and is asked to correctly read the entire word verbally or verbally produce each letter sound. For example is the word is “wuj,” the student could say “/w/ /u/ /j/” or “wuj.” Each correctly produced letter sound, whether individually or within a whole word, is scored as one point. The student is given one minute to produce as many sounds as possible and the total correct within one minute is the final score. Reliability estimates for D-NWF average near .80 (Elliot et al., 2001). The predictive validity of kindergarten D-NWF scores on 1st grade Woodcock-Johnson
Revised Reading readiness cluster scores is .59 (Good et al., 2001), and the criterion-related validity of D-NWF with DIBELS Oral Reading Fluency is .82 (Good & Kaminski, 1996).

*AIMSWEB Letter Sound Fluency (A-LSF).* A-LSF (Shinn & Shinn, 2002) measures a child’s ability to name individual phonemes that correspond to written letters (See Appendix E). For A-LSF, students are asked to name the sounds made by letters that are printed on a page. Both consonants and vowels are listed, but only short vowel sounds are scored as correct. The number of sounds the student speaks correctly in one minute is the final score. Criterion-related validity of LSF is .58 with the Woodcock Johnson III Broad Reading Cluster (Woodcock & Johnson, 1990) and .68 with the Test of Phonological Awareness Skills (Wagner et al., 1999) (Elliot, et al., 2001). Unlike the ability to name letters (D-LNF) letter-sound knowledge is required for fluent decoding of words, and this may explain its strength as a predictor of later reading success (Shinn & Shinn, 2002).

*Criterion Measures*

*DIBELS Oral Reading Fluency (D-ORF)* (see Appendix F for an example). D-ORF, like CBM-ORF, measures students’ ability to recognize words accurately and fluently within connected text (Kaminski & Good, 1996). Students are presented with printed passages and asked to read aloud for one minute. The total number of correctly read words in one minute is counted, and the median of three D-ORF scores administered is the final score used for decision-making. In one of the earliest efforts to examine the technical characteristics of ORF, Tindal, Marston, and Deno, (1983), conducted a comprehensive examination of the reliability of ORF. The study resulted in test-retest reliability coefficients ranging from .92 to .97, and equivalent forms reliability coefficients ranging from .89 to .94. Marston and Magnusson (1985), in another early study examined ORF for 309 students in grades one through three. Students’ reading
performance was assessed initially with ORF and several commercial assessments including SRA Vocabulary and SRA Comprehension. Reading instruction was delivered for 16 weeks, after which the students were assessed again with the same measures. ORF was found to be more sensitive to growth over time than either SRA Vocabulary or Comprehension. Teachers also reported satisfaction with the ORF measure after using it during the four-month experiment, despite some initial reservations. Since these early studies, evidence of the strong technical adequacy of ORF (supporting both CBM-ORF and D-ORF) has accumulated. As a measure of overall reading competence, the validity of ORF is well established (e.g., Deno, Mirkin, & Chiang, 1982; Espin, & Fogen, 1996; Fewster & MacMillan 2002; Fuchs & Fuchs, 1999; Fuchs et al., 2001; Fuchs, Fuchs, & Maxwell, 1988; Kame’enui et al., 2001; Markell & Deno, 1997; Marston, & Magnusson, 1985; Shinn & Good, 1992).

Within the DIBELS model, D-ORF is not indicated for use until winter of first-grade. Nevertheless, D-ORF was used as a criterion measure in the present study and administered in the spring of kindergarten. Published norm-referenced tests are more commonly used as criterion measures for validation studies than fluency-based measures like D-ORF. However, Speece et al. (2003) found that D-NWF and D-LNF administered in kindergarten contributed no significant variance to the prediction of scores on the WJ-R Letter-Word Identification and Word Attack (Woodcock & Johnson, 1990) in first grade. The implication is that norm-referenced tests may not be sufficiently sensitive to identify the full range of at-risk readers. Measures like the WJ-R are effective at differentiating between children across a wide range of ages, but subtests typically have a more limited representation of discrete age-appropriate reading skills within the items (Speece et al., 2003). Performance on D-ORF represents the integration of a range of necessary reading skills (Fuchs et al., 2001; Good et al., 2001), Thus D-ORF should be more
sensitive to critical skill deficits that may not be detected by norm-referenced diagnostic tests (Shinn & Good, 1992).

Test of Early Reading Ability-Third Edition (TERA-3). The TERA-3 is a norm-referenced measure of early reading ability. Instead of measuring skills associated with children’s readiness for learning to read (e.g., phonological awareness, rapid letter-naming), the TERA-3 targets skills that are more closely related to actual reading. Reid, et al., (2001) describes familiarity with the alphabet, understanding conventions of printed material, and finding meaning in printed material as emergent literacy skills. These skills are not conceptualized as precursors of reading, but as the beginning of literacy itself (Reid et al., 2001). For the proposed study, the TERA-3 is particularly well suited for use as a criterion measure since most of the DIBS that will be administered (e.g., D-LNF, A-LSF, and D-NWF) measure skills that involve printed material.

The TERA-3 has three subtests: Alphabet, Conventions, and Meaning. Alphabet includes items that test students’ ability to recognize print, name letters, and identify sight vocabulary. Conventions includes items measuring abilities such as book handling, distinguishing between upper and lower-case letters, capitalization, and spelling. Meaning measures an individual’s ability to gain meaning from printed material such as labels and logos, relating sentences to pictures, predicting text, and paraphrasing. Average coefficient alphas for subtests administered across ages were Alphabet (.90), Conventions (.83), and Meaning (.90) (Reid et al., 2001). Alternate form reliability coefficients for each subtest are reported by Reid, et al. (2001). All fall within the .82 to .95 range, while test-retest reliability coefficients range slightly higher, from .88 to .99. Reid, et al., (2001) reported considerable evidence supporting the validity of TERA-3 scores. For example, scores from the Stanford Achievement Tests Series Ninth Edition (SAT-9) (Harcourt, Brace, & Company, 1996) subtest of Total Reading, correlated with Alphabet (.62),
Conventions (.66), and with Meaning (.57). Likewise, scores from the Woodcock Reading Mastery Test-Revised (WRMT-R) Word Identification subtest (Woodcock, 1998) correlated with Alphabet (.61), Conventions (.47), and Meaning (.48). Reading Quotient scores of the WRMT-R correlated with Alphabet (.62), Conventions (.57), and Meaning (.61). All correlations are significant at the .01 level and support the concurrent validity of the TERA-3 with other validated measures or reading. In order to index overall performance on the TERA-3, a sum of the raw scores of each subtest was calculated, and this composite raw was used in the correlation and regression analyses. Since comparisons of performance between subtests were not conducted, standardized scores were not generated for the purposes of this study. Presented in a separate correlation matrix (Table 9), predictive and concurrent validity coefficients were calculated for DIBS at all benchmark periods with TERA-3 Alphabet, Conventions, and Meaning subtests.

**Procedures**

**Data Collection**

Students were individually administered the measures of early literacy skills examined in this study (see Table 5 for a summary of the administration schedule) three times. In the fall, students were administered the Initial Sound Fluency (D-ISF) and Letter-Naming Fluency (D-LNF) subtests of the DIBELS and the AimsWeb Letter-Sound Fluency (A-LSF) measure. In winter, Phoneme Segmentation Fluency (D-PSF) and Nonsense-Word Fluency (D-NWF) subtests are added to the administration. In the spring, the same DIBELS subtests (D-LNF, D-PSF, D-NWF) and the A-LSF measure are administered for a third time. In the spring, D-ORF and the TERA-3 were administered as criterion measures of reading. Two weeks after the fall testing session, a randomly selected sub-set of 50 students were administered the fall DIBS (D-
ISF, D-LNF, A-LSF) for a second time. These scores were correlated with the students’ original scores to calculate a coefficient of stability. Students from the University of Georgia assisted the school in the testing at each session. All were trained to administer the instruments by experienced faculty and graduate assistants, and inter-rater reliability was established (above 90% agreement) before testing at the school. During testing sessions throughout the year, inter-rater reliability checks were conducted for a minimum of 10 percent of administrations. Two trained assessors, one administering the measures and the other collecting reliability data, scored each measure independently. Reliability checks were balanced across the assessors at each benchmark.

Five to six students at a time were assessed in a reserved classroom with partitioned testing areas to minimize distractions. Students were pulled from classrooms for approximately fifteen minutes for each testing session. Teachers were made aware of testing schedules and assessors attempted to minimize the disruption to classes caused by testing sessions. Students who were not comfortable with testing or were reluctant to cooperate were not required to do so and incurred no negative consequences as a result. If students began testing and later choose to stop, they were returned to their teachers, again without penalty of any sort. Students who participated in testing were offered a small reward upon completion (e.g., stickers, small toys). For students who did not participate or who did not complete testing, the same reward was given to their teachers who made it available to them.

Data Analysis

Reliability and Predictive and Concurrent Criterion-Related Validity. Correlation, bivariate regression, and hierarchical regression were used to examine the predictive and concurrent validity of DIBS from each of the three benchmark assessment sessions. An
The intercorrelation matrix of all measures from the three benchmark sessions was generated. The strength of correlations between DIBS administered at fall and winter benchmarks with DIBS at later benchmarks were examined. Correlations of DIBS at fall, winter, and spring benchmarks with criterion measures administered in spring were also examined. A series of bivariate regressions were used to examine the amount of variance in each criterion measure that was explained independently by DIBS from each benchmark. Hierarchical regression models were constructed using DIBELS subtests and A-LSF from each benchmark to predict criterion measures. The order of entry into the regression equation was based on the established sequence of DIBELS subtest administration and the conceptualization of a continuum component early literacy skills that build upon one another, culminating in actual reading (Good et al., 2002). The order of entry of DIBELS and A-LSF into regression equations was based on this perspective: D-ISF and D-PSF were entered first. A-LSF was entered before D-NWF: Although both measure letter-sound fluency, A-LSF is thought to be less complex than D-NWF and therefore would be acquired first. D-LNF is considered an indicator of risk rather than a relevant skill that is developed on the path to literacy (Good et al., 2001). The strong relationship to later reading is likely due to the associations of D-LNF with other necessary reading skills that may be measured by the DIBS in the study. Therefore, D-LNF is entered into the regression equations last.

The final regression analyses were intended to identify the measures at each benchmark that were the strongest predictors of criterion measures. Although the theoretical framework that guided the previous set of regressions is well-supported, previous studies (e.g., Burke et al., in review; Elliot et al., 2001; Hintze et al., 2003; VanDerHeyden et al., 2001) suggest that the predictive strength of these measures may not be proportional to their presumed importance.
Thus, an empirical approach was used to determine whether differences in predictive strength were found among the measures when the influence of theory was removed. A blockwise selection procedure described by Pedhazur (1997) was used to determine the best set of predictors at each benchmark, based solely on the statistical significance of explained variance among the predictor variables.
CHAPTER 4

Results

Descriptive Statistics

The descriptive statistics for the fall, winter, and spring DIBS, along with the criterion measures in the spring are summarized in Table 7. Values for skewness and kurtosis all fall below an absolute value of 2, indicating that data are normally distributed. A test of the multivariate kurtosis resulted in an acceptable value of 1.097, indicating that data were multivariate-normally distributed. Together, these three indices suggest that data are suitable for further correlational analyses. Mean scores for DIBELS subtests at each benchmark were compared to the recommended cut score values delineating categories of risk (Good et al., 2001) (see Table 3). Categories correspond to the risk of students failing to perform adequately on the same subtest at the next benchmark administration. Mean scores for the sample on fall D-ISF, D-LNF, winter D-LNF, D-NWF, and spring D-LNF, and D-NWF were at or above the cut-score for the “low-risk” category. Mean scores on winter D-ISF, D-PSF, and spring D-PSF were within the range of “some risk”, but only slightly below the cutoff for low-risk. Mean scores of A-LSF were comparable to benchmark mean scores of a national sample reported by AimsWeb. Overall, scores of the sample across dependent measures were within the average range or only slightly below on selected measures.

Mean scores on D-ORF could not be compared to means of the DIBELS norm group, which are based on middle of first-grade administration of D-ORF. AimsWeb uses a standardized version of CBM-ORF at the beginning of first-grade, with technical characteristics
similar to D-ORF. The norm-group mean score of CBM-ORF for the beginning of first-grade was 7 words per minute and the mean D-ORF for the sample at the end of kindergarten was 7.5 words per minute. The scores on the TERA-3 composite were calculated by summing the raw scores on each subtest. Norms for this total are not included in the examiner’s manual for the TERA-3. Instead, the raw score for each subtest falling at the 50th percentile for students aged 5 years 6 months to 5 years 8 months was identified and summed: The sum of mean norm-group raw scores for Alphabet (14), Conventions (10) and Meaning (10) was 34, and the mean total raw score for the study sample was 39. The sample scores on both criterion measures of reading were above the values estimated to reflect the average performance of the population. Overall, the descriptive statistics suggest that the performance of the sample group is representative of the general population of same-aged children.

Reliability Estimates

Inter-rater and test-retest reliability were estimated using score to score correlations. Two weeks following the administration of DIBS in the fall (D-ISF, D-PSF, and A-LSF), each measure was re-administered to 50 students selected at random from the original sample. Coefficients of stability for each measure were computed using score to score correlations and are summarized in Table 6. Test-retest reliability for fall D-ISF was .72, fall D-LNF was .91, and fall A-LSF was .82. Inter-rater reliability coefficients ranged from .85 to .99.

Most evaluations of test reliability will include computing coefficients of internal consistency that estimate the homogeneity of item content. However, fluency-based measures like those that CBM and DIBS are scored based on correct items per unit of time; therefore, the items on individual tests are designed to be of equivalent difficulty and represent very similar content. Calculating coefficient alphas for the DIBS in this study would result in spuriously high
values since the item content is intended to be very homogenous. For this reason, internal consistency was not evaluated.

*Validity Coefficients*

Predictive and concurrent validity coefficients were calculated for individual DIBS with criterion measures administered at the end of kindergarten. Predictive validity coefficients for DIBS from fall and winter with DIBS at subsequent benchmarks were also calculated. Table 8 contains the inter-correlation matrix for all of the measures. All correlations were found to be significant at the .01 level except fall D-ISF with the TERA-3, which was significant at the .05 level.

*Predictive Criterion-Related Validity Coefficients*

Predictive validity coefficients for fall D-ISF with all DIBS in winter and spring and with both criterion measures were low; the largest coefficient was .34 with D-NWF in winter. Predictive validity coefficients for fall D-LNF were mostly moderate with winter and spring DIBS, ranging from .32 with D-PSF in spring to .61 with D-LNF in winter. Predictive validity coefficients for fall D-LNF was .55 for ORF and .44 for the TERA-3. Predictive validity coefficients for A-LSF in fall winter and spring DIBS were also mostly moderate, ranging from .24 with D-ISF in fall to .48 with both D-NWF and A-LSF in winter, and the lowest predictive validity coefficient was .24 with D-ISF in winter. Predictive validity coefficients for fall A-LSF was .44 with ORF and .42 with the TERA-3. Intercorrelations of the fall DIBS show D-LNF and A-LSF correlate at .64, and D-ISF correlates with D-LNF and A-LSF at .28 and .29 respectively.

Predictive validity coefficients for winter DIBS with spring DIBS and the criterion measures were generally higher than fall DIBS. Again, predictive validity coefficients for D-ISF in winter are the weakest of the set ranging from .23 with spring A-LSF to .32 with both D-PSF.
and D-NWF in spring. Predictive validity coefficients for winter D-ISF were .29 with D-ORF
and .30 with the TERA-3. Winter D-LNF had moderate predictive validity coefficients with
spring D-LNF, D-NWF, and A-LSF (.66, .46, and .54 respectively); and with D-ORF and
TERA-3 (.58 and .55). Predictive validity coefficients for winter D-PSF was .43 for D-ORF .47
for the TERA-3. Winter A-LSF had strong predictive validity coefficients with spring DIBS
with the exception of the predictive validity coefficient of .30 with spring D-PSF. Predictive
validity coefficients for winter A-LSF was .66 with ORF and .58 with the TERA-3. In sum, the
strongest predictive validity coefficients of winter DIBS with ORF and the TERA-3 were those
of A-LSF, D-NWF and D-LNF. Intercorrelations of winter DIBS resulted in strong coefficients
for winter D-LNF with winter A-LSF (.73) and with winter D-NWF (.71). An even stronger
correlation coefficient of .81 is found between winter D-NWF and winter A-LSF. Correlations of
winter D-ISF with other winter DIBS resulted in mostly weak coefficients; the highest was .38
with winter A-LSF.

**Concurrent Criterion-Related Validity Coefficients**

Concurrent validity coefficients were calculated for spring DIBS with D-ORF and the
TERA-3. Concurrent validity coefficients for spring D-NWF, D-LNF, and A-LSF with D-ORF
were .66, .64, and .53 respectively. The concurrent validity coefficient of spring D-PSF with D-
ORF (.35) was considerably lower. The strength of concurrent validity coefficients for spring
DIBS with the TERA-3 followed an identical pattern as those with D-ORF: Concurrent validity
coefficients of spring D-NWF, D-LNF, and A-LSF with the TERA-3 were .51, .49, and .42
respectively. Again, the lowest concurrent validity coefficient was .35 for D-PSF with the
TERA-3. Intercorrelation coefficients of spring DIBS were similar in relative strength to those
observed in winter. Inter-correlations of print-related measures, D-LNF, D-NWF, and A-LSF,
were moderate to strong, while correlations of winter D-PSF with other winter DIBS were weak to moderate.

Predictive and Concurrent Validity Coefficients of DIBS with TERA-3 Subtests. A correlation matrix of fall, winter, and spring DIBS with TERA-3 Alphabet, Conventions, and meaning subtests is presented in Table 9. Except for the coefficient for fall D-ISF with TERA-3 Conventions ($r = .03$, ns), all coefficients were significant. In fall, D-LNF and A-LSF correlated modestly with all three subtests. Fall D-LNF correlated strongest with TERA-3 Meaning (.36) and fall D-LNF correlated strongest with TERA-3 Alphabet (.38). Fall D-ISF correlated weakly with all three TERA-3 subtests. In winter, D-LNF and A-LSF coefficients increase, reflecting relationships that are more moderate with TERA-3 subtests. Winter D-NWF had similarly moderate correlations; the coefficient of winter D-NWF with TERA-3 Alphabet was .52, and the coefficient with TERA-3 Conventions was .46—the highest of the four winter DIBS. Coefficients for winter D-ISF are all significant, but remain weak, ranging from .19 with TERA-3 Meaning to .25 with TERA-3 Conventions. Winter D-PSF correlates modestly with all three TERA-3 subtests; the strongest coefficient is .38 with TERA-3 Alphabet. In spring, coefficients of DIBS with TERA-3 subtests were all slightly lower than those from winter. As in winter, D-LNF from spring, again had the strongest correlation to TERA-3 Alphabet (.51) of the four spring DIBS. The coefficient of spring D-NWF with TERA-3 Conventions (.44) was the strongest of the spring DIBS. Finally, coefficient of spring D-LNF with TERA-3 Meaning (.30) was the strongest of the four. In general, the DIBS involving print and requiring alphabetic understanding (D-LNF, A-LSF, and D-NWF) showed consistently stronger correlations with the three TERA-3 subtests at all benchmarks than did the measures requiring phonological
awareness skills (D-ISF and D-PSF). As a group, winter DIBS correlated slightly higher with TERA-3 subtests than did the same group of DIBS in spring.

Regression Analyses

Three sets of regression analyses were used for both predictive and concurrent analyses. The results of the predictive analyses of reading for fall and winter DIBS are presented first, followed by the concurrent analyses of reading for spring DIBS. The results of predictive analyses for fall DIBS are summarized in Table 10, the results for winter DIBS are summarized in Table 11, and the concurrent analyses for spring DIBS are summarized in Table 12. In the tables and in the following section, values for the adjusted \( R^2 \) are reported, as are standardized Beta coefficients. Each table presents three sets of regression analyses that were conducted for the groups of DIBS at each benchmark and for each of the criterion measures. First, to determine the overall variance in D-ORF and the TERA-3 (composite score) that was explained by individual DIBS, simple bivariate regression analyses were conducted. Second, hierarchical regression analyses were conducted to examine the amount of variance in criterion measures explained by DIBS when the order of entry into the regression equation was based on a specific theoretical perspective of reading acquisition, consistent with the perspective that underlies the sequencing of DIBELS (Good, Kaminski, Smith, Simmons, Kame’enui, & Wallin, 2003). Measures targeting phonological awareness (D-ISF & D-PSF) were entered first, followed by measures targeting alphabetic understanding. A-LSF was entered into the equation before D-NWF because A-LSF was presumed to measure less complex skills that should logically precede D-NWF, which was entered next. D-LNF was entered into the equation last because it is conceptualized as an indicator of risk (for reading failure) rather than a measure of necessary reading skills. The relative amount of variance in criterion measures explained by each
successive model in the hierarchical regression analyses was determined. In contrast to the hierarchical regression analyses, the third set of predictive and concurrent analyses was conducted from an empirical, atheoretical perspective. The goal was not to make causal inferences regarding DIBS and reading, but was rather to determine the optimal set of DIBS at each benchmark that could be used to predict reading in spring. Using the forward blockwise selection procedure, described by Pedhazur (1997), a set of DIBS from one benchmark was entered into the regression equation simultaneously as a block. The measures found to be significant were then entered into successive hierarchical regression equations; the non-significant measures were removed. Successive hierarchical regressions were conducted with the order of the significant variables in the subset reversed. The amount of significant unique variance in criterion measures explained by each measure was determined.

**Predictive Regression Analyses**

*Predicting Reading with DIBS Administered in Fall.* To determine the amount of variance in D-ORF and the composite score of the TERA-3 was explained by each fall DIBS measure, bivariate regression analyses were conducted. Individually, the three fall DIBS each explained a significant (\(p < .01\)) amounts of variance in D-ORF: D-ISF explained 6%, D-LNF explained 29%, and A-LSF explained 19% (Table 10, Models 1, 2, & 3). The amount of variance in the TERA-3 explained by D-ISF was 1% (\(p < .05\)); A-LSF explained 17% (\(p < .01\)); and D-LNF explained 19% (\(p < .01\)).

For the theoretically-driven hierarchical regression (Table 10, Model 4), D-ISF was entered first, followed by A-LSF and D-LNF. In the models predicting D-ORF, D-ISF accounted for 6% (\(p < .01\)) of the variance, the addition of A-LSF explained an additional 24% (\(p < .01\)), and the final addition of D-LNF explained only 1% more (\(p=ns\)). The results of the hierarchical
model explaining the TERA-3 (Table 10, Model 4) resulted in D-ISF explaining 1%, ($p < .01$); A-LSF explaining an additional 17%, ($p < .01$); and D-LNF adding 3%, ($p < .01$).

In the last set of regression analyses, D-ISF, A-LSF, and D-LNF were entered into the regression equation simultaneously (Table 10, Model 5). For predicting D-ORF, only D-LNF was found to account for significant variance ($R^2 = .29, p < .01$). When D-LNF was entered first, neither D-ISF nor A-LSF explained additional significant variance in D-ORF. The same initial block was entered into another regression analysis to predict TERA-3 (Table 11, Model 6). The model explained 21% of variance in TERA-3 ($R^2 = .21, p < .01$); D-LNF and A-LSF were both significant, while D-ISF was not. D-LNF and A-LSF, in this order, were then entered into a second regression analysis. D-LNF accounted for 19% ($p < .01$) of variance in TERA-3, and A-LSF accounted for 3% ($p < .01$). These variables were entered again in the reverse order. When entered first, A-LSF accounted for 17% ($p < .01$) and D-LNF accounted for an additional 5% ($p < .01$).

*Predicting Reading with DIBS Administered in Winter.* The same series of regression analyses was conducted with winter DIBS to explain variance in D-ORF and TERA-3, and the results are summarized in Table 11. The bivariate regression analyses were used to explain the amount of variance in D-ORF and TERA-3 explained by winter DIBS individually. For prediction of D-ORF, winter D-ISF accounted for 8%, ($p < .01$); D-PSF accounted for 18%, ($p < .01$); D-NWF accounted for 48%, ($p < .01$); and A-LSF accounted for 42%, ($p < .01$). For prediction of TERA-3, winter D-ISF accounted for 8%, ($p < .01$); D-PSF for 21%, ($p < .01$); D-NWF for 35%, ($p < .01$); and A-LSF for 33%, ($p < .01$).

Results of the theoretically-driven hierarchical regression analyses are presented in Table 11, Model 6. Winter DIBS were entered one at a time into the regression analysis in the
following order: D-ISF, D-PSF, A-LSF, D-NWF, and D-LNF. The entire model accounted for 50% or the variance in D-ORF. After D-ISF was entered ($R^2 = .08$), D-PSF accounted for 11%, ($p < .01$) additional variance. A-LSF added 24%, ($p < .01$). D-NWF added 6%, ($p < .01$). Finally, D-LNF added less than 1% (ns) of explained variance in D-ORF. The same model was then used to predict TERA-3, and accounted for a total of 40% of the variance ($R^2 = .40$, $p < .01$). As entered, D-ISF accounted for 8%, ($p < .01$); D-PSF added 14%, ($p < .01$); A-LSF added 14%, ($p < .01$); D-NWF added 2%, ($p < .05$); and D-LNF added 1%, ($p < .05$).

In the final set of empirically-driven analyses, all DIBS were entered into a regression analysis simultaneously as a block (Table 11, Model 7). For the analysis predicting D-ORF, only D-NWF and A-LSF were significant. D-NWF and A-LSF were then entered into a second analysis in that order. D-NWF accounted for 48%, ($p < .01$) of the variance on D-ORF, and A-LSF accounted for an additional 2%, ($p < .01$). In the reverse order, A-LSF accounted for 42% ($p < .01$), and D-NWF accounted for an additional 7%, ($p < .01$).

Using the same procedure to predict TERA-3 (Table 12, Model 8), the block of four DIBS was entered simultaneously. D-NWF, and D-LNF were significant at the level of $p < .01$, and D-PSF was significant at $p < .05$. Together, these three DIBS accounted for 50% of the variance in TERA-3 ($R^2 = .50$, $p < .01$). D-ISF and A-LSF did not account for a significant amount of variance in TERA-3. D-NWF, D-LNF, and D-PSF were then entered into a second hierarchical regression analysis with D-NWF entered first based on the strength of the standardized beta coefficient (beta = .314, $p < .01$), followed by D-LNF and D-PSF. D-NWF accounted for 35%, ($p < .01$) of the variance in TERA-3, D-LNF added 3%, ($p < .01$) additional variance, and D-PSF added a final 1%, ($p < .05$). In a third analysis, the order of D-LNF and D-
PSF was reversed. To the 35% accounted for by D-NWF, D-PSF added 1%, \( p < .05 \) and D-LNF added 3%, \( p < .01 \).

**Concurrent Criterion-Related Validity**

*Concurrent analyses of spring DIBS.* The series of regression analyses used to evaluate the predictive validity of fall and winter DIBS was used to evaluate the concurrent criterion-related validity of spring DIBS with D-ORF and TERA-3. The results of criterion analyses are presented in Table 12. Bivariate regressions were conducted using each DIBS in spring as a predictor of D-ORF and TERA-3 (Table 12, Models 1 through 4). D-PSF explained 9%, \( p < .01 \) of the variance in D-ORF; D-NWF explained 43%, \( p < .01 \); D-LNF explained 40%, \( p < .01 \); and A-LSF explained 27%, \( p < .01 \). For the analyses predicting TERA-3, D-PSF explained 11% of the variance, \( p < .01 \); D-NWF explained 25%, \( p < .01 \); D-LNF explained 23%, \( p < .01 \) and A-LSF explained 17%, \( p < .01 \).

For the theoretically-driven hierarchical analysis of spring DIBS predicting D-ORF and TERA-3, DIBS were entered into regression analyses in the following order: D-PSF, A-LSF, D-NWF, D-LNF (Table 12, Model 5). For the analyses predicting D-ORF, D-PSF was entered first \( (R^2=.09, p < .01) \) followed by A-LSF, that accounted for an additional 19% of the variance, \( p < .01 \); D-NWF added 15%, \( p < .01 \); and D-LNF added 8%, \( p < .01 \). The total variance in D-ORF explained by the model was 52% \( (R^2=.52, p < .01) \). Using the same order of entry to predict the TERA-3, D-PSF accounted for 11% \( ( p < .01) \) of the variance; A-LSF added 9%, \( p < .01 \); D-NWF added 6%, \( p < .01 \); and D-LNF added 4%, \( p < .05 \). The total variance in TERA-3 explained by the model was 31% \( (R^2=.32, p < .01) \).

For the empirically-driven analyses of spring DIBS predicting D-ORF and TERA-3, all DIBS were entered into a regression analysis simultaneously (Table 12, Model 6). In the analysis
predicting D-ORF, D-LNF and D-NWF accounted for significant variance ($R^2=.52, p < .01$). D-PSF and A-LSF did not account for significant variance in D-ORF. D-NWF and D-LNF were then entered into a second regression analysis in that order. D-NWF independently explained 42%, ($p < .01$) of the variance in D-ORF. Adding D-LNF to the equation explained 9% more variance, ($p < .01$). In the reverse order, D-LNF explained 40% of the variance in D-ORF and D-NWF added 12%, ($p < .01$).

In the empirically-driven regression analysis using spring DIBS to predict TERA-3 (Table 12, Model 6), simultaneous entry of D-PSF, A-LSF, D-NWF, and D-LNF resulted in D-NWF and D-LNF accounting for significant variance in TERA-3 ($R^2=.31, p < .01$), and D-PSF and A-LSF accounting for no significant variance. D-NWF and D-LNF were entered into a hierarchical regression in that order. D-NWF accounted for 26%, ($p < .01$) of the variance in TERA-3, and the addition of D-LNF accounted for 5% more variance, ($p < .01$). In the reverse order, D-LNF accounted for 23%, ($p < .01$), and D-NWF accounted for 7%, ($p < .01$).
CHAPTER 5

Discussion

In light of the accumulating evidence that reading difficulties can be prevented with effective early intervention (Adams, 1990; Torgesen & Wagner, 2002), identifying valid measures for predicting reading failure is imperative. Dynamic indicators of basic skills, or DIBS, have been used to identify students in need of early literacy intervention before poor reading trajectories are established (Good et al., 2001, 2002). The use of DIBS in early literacy within a preventative approach to reading failure is supported by a growing body of evidence, but many questions remain unresolved, particularly regarding which measures are most valid for predicting reading failure in kindergarten and when in kindergarten is the best time to measure. Thus, the purpose of this study was to assess the validity of a set of DIBS in early literacy administered in kindergarten for the purpose of identifying students at-risk of reading failure. The measures of interest were DIBELS (Good & Kaminski, 1996) subtests indicated for kindergarten and AimsWeb Letter-Sound Fluency (Shinn & Shinn, 2002). Extending the results of previous evaluations of validity, criterion measures of reading were chosen for this study that represented a broad range of reading skills thought to be critically necessary at the end of kindergarten (Adams, 1990; National Reading Panel, 2000; National Research Council, 1998; Good & Kaminski, 2002; Hammill, 2004). Concurrent and predictive criterion-related validity coefficients were evaluated for sets of DIBS administered at fall, winter, and spring benchmarks during the school year. At each benchmark, the contribution of individual DIBS to criterion
measures of reading was determined, and the strongest predictors of reading outcomes were identified.

The results of predictive and concurrent analyses were consistent with previous research demonstrating that DIBELS Letter-Naming Fluency and DIBELS Nonsense-Word Fluency are the most consistent predictors of reading outcomes when administered at the middle and end of kindergarten (Elliot et al., 2001; Schatschnieder et al., 2004; Speece et al., 2003; VanDerHeyden et al., 2001). Initial support was also demonstrated for the validity of AimsWeb Letter-Sound Fluency as a potentially useful indicator of reading when administered in the winter of kindergarten. The results are discussed in more detail within the context of the research questions.

*Question 1.* Will scores from identical forms of DIBS (DIBELS subtests and A-LSF) administered two weeks apart correlate strongly (r > .80), demonstrating that the skills measured are stable?

The test-retest reliability coefficients calculated were limited to the subtests administered in the fall: D-ISF, D-LNF and A-LSF and were administered to a randomly selected subset of 50 students, two weeks after the school-wide assessment at the fall benchmark. Coefficients for D-LNF and A-LSF were both above .80, indicating an acceptable level of stability (Salvia and Ysseldyke, 2004). These values are consistent with reliability coefficients found by Elliot et al. (2001) and Good et al. (2001). A lower coefficient was found for D-ISF (r = .72), but this value is also consistent with the coefficient found by Good, et al., (2001). This somewhat lower value due to the nature of the test administration requiring assessors to time the latency between the presentation of a picture stimulus and a students’ response. The difficulty of rapid repeated timings may contribute to D-ISF scoring errors.
Inter-rater reliability coefficients were calculated in winter and spring based on score to score correlations of simultaneous scoring of DIBS. Values were all above .90 except the coefficients for D-ISF that were somewhat lower. All are within the acceptable range for the intended purposes of testing (Salvia & Ysseldyke, 2004); However, the total $n$ for inter-rater reliability estimates was low (below 20 percent of total $n$) in winter ($n = 22$) and spring ($n = 15$). The inter-rater coefficients for these benchmarks should be interpreted with caution.

**Question 2. What are the correlations between DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF), at the middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF), and at the end (D-LNF, D-PSF, D-NWF, and A-LSF) of kindergarten?**

**Fall DIBS.** Evaluating the predictive validity of DIBS in early literacy used within an outcomes-based approach requires not only examining associations with criterion measures of reading at a later date, but also the associations between DIBS at one benchmark and DIBS at later benchmarks. Performance on simpler skills is presumed to predict performance on later more complex skills that lie on a continuum leading to reading acquisition (Good et al., 2003). The correlations found between fall DIBS and winter DIBS were all statistically significant, with fall D-LNF having the highest correlations with winter DIBS. The additional measure of A-LSF correlated with all winter DIBS more strongly than did fall D-ISF. Although, based on previous evidence, a somewhat weaker relationship between D-ISF and later DIBS was expected (Burke et al., in review; Good et al., 2001), that fall A-LSF was more predictive of winter DIBS than fall D-ISF was surprising considering the importance of early phonological awareness skills suggested in the literature. These results support the findings of Elliot et al. (2001) and VanDerHeyden et al., (2001) who found generally stronger relationships between measures of letter-sound fluency and reading than initial-sound fluency and reading. Perhaps the most
interesting finding is that fall A-LSF correlated with winter D-PSF at .40, while fall D-ISF correlated with winter D-PSF at only .33. This suggests that sound-symbol correspondence tasks like A-LSF may be useful predictors of reading when administered at the beginning of kindergarten. A similar pattern of correlations between fall DIBS and spring DIBS was evident: Fall D-LNF had the strongest correlations with spring DIBS. Fall A-LSF again correlated moderately with all spring DIBS, and all correlations were stronger that those between fall D-ISF and spring DIBS. This pattern provides some confirmation that A-LSF administered in the fall is a valid predictor of important literacy skills. The values of the correlations between fall D-ISF and later DIBS are relatively consistent with earlier studies, however, the coefficients remain generally weak. Stronger support is demonstrated for the use of fall A-LSF as a predictor of later reading skills.

Winter DIBS. On the continuum of reading skill acquisition proposed by Good et al. (2003), proficiency on winter D-PSF is a critical indicator of reading progress, and the development of phonemic awareness skills in the middle of kindergarten supports performance on D-NWF in the spring. Consistent with this logic, winter D-PSF was found to correlate (.46) with spring D-NWF, the highest correlation between winter D-PSF and any spring measure. Correlations between winter D-ISF and spring measures were consistently weak, while winter D-NWF, of all of the winter DIBS, correlated most strongly with spring measures. Both winter D-LNF and A-LSF correlated moderately with spring measures, and nearly all of the coefficients were higher than those from similar correlations of fall D-LNF and A-LSF with winter DIBS. This increase suggests that a developmental or instructional influence may be present that raises the correlations of these measures with reading as the kindergarten year progresses.
Question 3. What are the correlations between DIBS administered at the beginning (D-ISF, D-LNF, and A-LSF) and middle (D-ISF, D-PSF, D-NWF, D-LNF, and A-LSF) of kindergarten, and criterion measures of reading ability administered in the spring (TERA-3, D-ORF)?

Fall DIBS. An examination of the predictive validity coefficients for DIBS administered in the fall for predicting the criterion measures of reading found that D-LNF and A-LSF predicted D-ORF ($r=.55$ & $.44$ respectively) and TERA-3 ($r=.44$ & $.42$ respectively) better than fall D-ISF ($r=.27$ with D-ORF and $.15$ with TERA-3). These results provide further support for the predictive validity of fall D-LNF and A-LSF. Both coefficients between fall D-ISF and the outcome measures were statistically significant, but the weakness of the coefficients relative to those of D-LNF and A-LSF brings into question the practical significance of the findings.

Winter DIBS. Correlations of winter D-LNF and A-LSF with year-end criterion measures were moderate, ranging from $.55$ to $.56$. Correlations with criterion measures for both D-LNF and A-LSF increased from the values found in fall. This result suggests that D-LNF scores do not result in a ceiling effect in the winter of kindergarten as hypothesized. The moderate correlations found for winter D-PSF provide some support for predictive validity, but correlations with criterion measures were considerably weaker for D-PSF than for D-LNF, A-LSF, or D-NWF that all involve print. D-NWF correlated the strongest with both D-ORF ($.70$) and TERA-3 ($.60$) of all of the winter DIBS. Consistent with fall correlations of DIBS with criterion measures, the winter measures of phonological awareness skills (D-ISF and D-PSF) had lower correlations with criterion measures than measures involving print and letter-sound knowledge (D-LNF, A-LSF, and D-NWF). These results suggest that measures of skills with print in the middle of kindergarten are superior to measures of phonological awareness skills for predicting reading outcomes.
Question 4 & Question 5. Will particular DIBELS measures, of those administered at the beginning and at the middle of kindergarten, explain a unique amount of the variance in criterion measures of reading administered at the end of kindergarten (TERA-3 and D-ORF)? Will A-LSF fall and winter benchmarks explain significant unique variance in the criterion measures?

Fall Models. Questions 4 and 5 were investigated within the same series of regression analyses and are discussed together. To evaluate the predictive ability of fall and winter DIBS, a series of hierarchical regression analyses were conducted. The first series was theoretically-driven, based on the relative importance of early literacy skills indicated in the DIBELS outcomes-driven model (Good et al., 2001). The second series of regressions was empirically-driven, intended to identify the optimal set of predictors for each benchmark in the absence of theory. Simultaneous regressions were conducted to identify variables contributing significant variance to criterion measures. These sub-sets of variables were then entered into hierarchical regressions with the order of entry alternated. The amounts of unique variance explained by each variable were determined. In the fall, the theory-driven hierarchical model explained 31% \((p < .01)\) of the overall variance in D-ORF and 21\% \((p < .01)\) in TERA-3. A-LSF was identified in both theoretical models as the variable explaining the majority of variance in the criterion measures \((R^2 = .24, p < .01, \text{for D-ORF}; \ R^2 = .17, p < .01 \text{ for TERA-3})\).

In contrast to the theoretical models, the results of the empirical models indicated that D-LNF was the most predictive of D-ORF and TERA-3 of the fall DIBS. D-LNF accounted for a 29\% of the variance in D-ORF and A-LSF did not explain additional significant variance. D-LNF explained the majority of unique variance in TERA-3 \((R^2 = .19, \ p < .01)\), and A-LSF explained an additional 3\% of the unique variance. In both the models predicting D-ORF and the
TERA-3, D-ISF did not account for significant amounts of variance. The results confirm the study hypotheses that fall D-LNF would account for the most unique variance in outcome measures. Furthermore, results support the role of D-LNF as a strong predictor of reading in the fall of kindergarten (Burke et al., in review; Elliot et al., 2001; Good et al., 2000, 2001; Speece et al., 2003). A-LSF in the fall did not explain unique variance in D-ORF, however, a small but significant percentage was explained in the TERA-3. Whether A-LSF would be useful as a predictor of reading in the fall is unclear. Letter-sound correspondence must be taught directly (Ehri, 1989). Students are not expected to have knowledge of letter-sounds in kindergarten. Therefore, measures of A-LSF at the beginning of kindergarten may provide an indication of students’ exposure to explicit instruction in letter-sound correspondences prior to kindergarten. The increase in mean scores of A-LSF in fall (4.4) to A-LSF in winter (20) observed in this study are clear evidence that students’ skills with identifying letter-sound correspondences improved in the first half of kindergarten, and this increase is almost certainly the result of instruction. A-LSF may become a better predictor of reading once instruction has begun, and may function as an early indicator of students’ responsiveness to instruction in kindergarten.

Winter DIBS. The predictive analyses of winter DIBS demonstrated the strength of DIBS involving skills with print to predict the criterion measures of reading. Independently, all of the DIBS in winter explained significant amounts of variance in both D-ORF and the TERA-3. Subsequent analyses revealed that much of this variance was shared. In the theoretical models predicting criterion measures with winter DIBS, A-LSF accounted for the most variance in D-ORF (24%), with the entire model accounting for 50%. Winter D-PSF and A-LSF each accounted for 14% of the overall variance ($R^2 = .40$) in the TERA-3. These results are inconsistent with a theoretical perspective placing greatest importance on phonological
awareness skills in the middle of kindergarten. In both models, measures involving print accounted for as much or more variance than phonological awareness measures.

The empirical models resulted in similar findings. Winter D-NWF and A-LSF accounted for 50% of the variance in D-ORF, with D-NWF independently accounting for 48%. Winter D-NWF was also the strongest predictor of the TERA-3, accounting for 35% of the variance. Winter D-LNF and D-PSF explained significant but small amounts of additional variance in the TERA-3. As expected, winter D-NWF accounted for the most unique variance in D-ORF and the TERA-3. These results support the work of Speece (2003) who demonstrated the utility of D-NWF at the end of kindergarten for predicting reading in first grade, and provides initial support for the use of D-NWF in winter of kindergarten to predict later of reading. Results showed similar proportions of explained variance are found for winter A-LSF. Winter A-LSF uniquely accounted for 5% of the variance in D-ORF, and shared 42% of explained variance in D-ORF with D-NWF. This overlap is consistent with the hypothesis that D-NWF and A-LSF measure similar skills, however, whether A-LSF would be uniquely useful as a predictor of reading in winter is unclear. A-LSF may have potential for addressing part of the problem with over-prediction in kindergarten (O’Connor & Jenkins, 1999). A-LSF requires skills that lie somewhat in-between letter-naming ability and D-NWF. Students must not only identify letters from print but must produce the corresponding phoneme symbolized by the letter. D-NWF presents phonemes in the context of phonemically regular nonsense words rather than letters in isolation. Students are also instructed to blend the phonemes together if they are able. Reasonably, students’ ability to identify individual phonemes from print fluently will precede the ability to blend three-phoneme words together. Measures like A-LSF may add another level of specificity to DIBELS measures in kindergarten, resulting in identification that is more accurate.
Question 6. What are the correlations between DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) and criterion measures of reading (TERA-3 and ORF)?

Concurrent validity coefficients for spring DIBS with D-ORF and the TERA-3 (Table 8) were calculated to evaluate the relationships between DIBS and presumed measures of similar constructs. Correlations between D-NWF and both criterion measures supported the hypothesis that D-NWF would have the strongest concurrent validity coefficients of the four DIBS administered in spring. D-LNF and A-LSF both correlated moderately with D-ORF and the TERA-3, providing yet another example of the stable association these measures have with reading outcomes throughout kindergarten. The correlations of spring D-PSF with D-ORF (.32) and the TERA-3 (.35) are lower than values for the other spring DIBS. These correlations are also lower than were found in previous studies that examined the relationship of D-PSF with D-ORF (Good et al., 2000, 2001; Kaminski & Good, 1996). Interestingly, D-PSF in winter correlated more strongly with criterion measures than did D-PSF in spring. Conversely, while correlations of DIBS involving print with criterion measures generally increased from winter to spring.

Question 7. Will certain DIBS administered in spring (D-LNF, D-PSF, D-NWF, and A-LSF) explain unique amounts of variance in criterion measures of reading (TERA-3 and ORF)?

The bivariate regression analyses resulted in each spring measure explaining significant variance in D-ORF and the TERA-3 (Table 12, Models 1-4). For both criterion measures, D-NWF and D-LNF accounted for the most variance individually. The theory-driven hierarchical models (Table 12, Model 5) explained 52% of the variance in D-ORF and 31% of the variance in the TERA-3. Within the models, A-LSF accounted for the most variance in D-ORF (19%) and D-PSF accounted for the most variance in the TERA-3 (9%).
The results of the empirically-driven models were quite different (Table 12, Models 6 & 7). Blockwise entry of spring DIBS identified D-NWF and D-LNF as the best predictors of both D-ORF and the TERA-3 were. Together, D-LNF and D-NWF explained 40% of the variance in D-ORF. Individually, D-NWF uniquely explained 12% of the variance, and D-LNF explained 9% of the variance. Similar results were found for predicting the TERA-3: Together the D-LNF and D-NWF accounted for all of the explained variance in the original model ($R^2 = .31, p < .01$). D-NWF accounted for 7% unique variance in the TERA-3 and D-LNF accounted for 5% unique variance. These results are very similar to the findings of Speece, et al., (2003) who used D-LNF and D-NWF at the end of kindergarten to predict scores on CBM-ORF (Speece & Case, 2001). D-LNF and D-NWF accounted for 34% of the variance in CBM-ORF administered in first grade, and both accounted for unique variance. The present study replicates these results, demonstrating the concurrent validity of D-NWF and D-LNF as well as providing further support for the use of D-ORF as an outcome measure of reading in the spring of kindergarten. The predictive strength D-NWF confirmed the hypothesis that D-NWF in spring would account for the most variance in the outcome measures. Surprising, however, was that D-LNF was more predictive than A-LSF or D-PSF in spring. The similarity of letter-naming tasks with letter-sound tasks has been proposed as an explanation of the predictive strength of letter-naming fluency measures (Ehri & Wilce, 1985; Scarborough, 1998). Following this logic, D-LNF would function as an indicator of letter-sound skills, and letter-sound skills are presumed to be more relevant than letter-naming to the process of reading acquisition. This hypothesis is not supported by the results: A-LSF, a measure of letter-sound knowledge, did not account for significant variance in the criterion measures in spring, while D-LNF did. The relationship of A-LSF to D-NWF could result in shared explained variance in the criterion measures, but this would not fully explain the effect of
D-LNF since it independently accounts for significant unique variance in the criterion measures. The hypothesis that D-LNF would result in a ceiling effect by spring of kindergarten was also not supported. Many children may have learned the alphabet by the end of kindergarten, but this study found that D-LNF still discriminated between children based on their performance on the outcome measures. One reason for this may be that the fluency component of D-LNF is the important predictive element. However, Schatschnieder et al. (2004) found letter-naming and RAN tasks to correlate strongly at the beginning of kindergarten, but less strongly by the end of kindergarten. That D-LNF in spring surpassed D-PSF in predictive strength was also unexpected. Whether phonemic awareness is a pre-requisite to acquiring alphabetic understanding (Torgesen, et al., 1994) or whether phonemic awareness develops as a result of learning to read (Ehri, 1989), the relationship of phonemic awareness to reading was expected to become more robust as students progressed through kindergarten (Schatschnieder et al., 2004). The observed relationships of D-PSF and the criterion measures do not support this logic.

Implications for Research and Practice

For the goal of identifying students in need of early literacy intervention, using DIBS in early literacy represents a promising approach that is supported by well-accepted theory and a growing body of empirical evidence. The present study investigated the validity of DIBS in early literacy used to predict reading performance during kindergarten. The results of the study suggest (a) DIBS measuring literacy skills involving knowledge of print, particularly D-NWF and D-LNF, when administered in kindergarten are strong predictors of year-end reading, (b) DIBS measuring phonological awareness administered in kindergarten (D-ISF and D-PSF) are less predictive of reading outcomes than measures involving skills with print, (c) D-LNF is an effective predictor of reading at the beginning of kindergarten despite presumed developmental
differences among children, and (d) A-LSF may be a useful addition to DIBELS measures administered in kindergarten. Several implications should be considered related to the use of DIBS in early literacy for identifying students in need of reading intervention in kindergarten.

First, the strong estimates of reliability and validity of D-LNF corroborate previous findings indicating letter-naming fluency tasks are the strongest single predictors of reading at the beginning of kindergarten (Speece et al., 2003). Letter-naming fluency has been found to over-predict students early in kindergarten (Elliot et al., 2001; O’Connor & Jenkins, 1999). However, until measures with greater predictive accuracy are identified, D-LNF remains the most viable predictor of reading difficulties when administered in the fall. Schatschnieder et al., (2004) found that D-LNF was unstable across the kindergarten year, and that the strength of correlations between D-LNF and reading tapered at the end of the year. The results of this study indicate that D-LNF was stable across the year, and at the end of kindergarten accounted for significant unique variance in both criterion measures. Practitioners who choose to use D-LNF for fall screening purposes should pay careful attention to winter scores and verify whether at-risk status is still demonstrated for students identified with D-LNF in the fall. For winter screenings and progress monitoring, D-NWF was the strongest predictor. Within the outcomes-driven model proposed by Good et al., (2001) D-NWF is considered an emerging skill in winter and is considered secondary in importance to D-PSF as an indicator of progress. Though the results of this study indicate that D-NWF may be useful as a predictor, replications of results are needed before changes in winter screening practices are advisable. In particular, predictive accuracy analyses, which were not conducted in the present study, would confirm whether D-NWF in winter accurately predicts reading performance in first and second grade in addition to the end of kindergarten.
Both D-ISF and D-PSF demonstrated weaker predictive and concurrent validity than did the other measures involving print. Based on these results, measures of phonological awareness should not be dismissed as potentially useful predictors of reading. The poor estimates of reliability found in this study for D-ISF suggest that aspects of the instrument may cause scoring errors. Assessors must juggle several tasks at once when administering D-ISF, and must make quick judgments regarding the accuracy of students’ verbal responses. D-ISF may be best administered by a school psychologist or other competent professional with experienced with test administration. Future research should investigate whether D-ISF administration can be modified to result in higher reliability estimates and potentially greater predictive validity. The results for D-PSF in winter and in spring are more difficult to explain. Mean scores on D-PSF in winter (16.65) and D-PSF in spring (32.62) clearly indicate growth in phonemic segmentation ability. Despite this, correlations with D-ORF and the TERA-3 diminished considerably from winter to spring. This evidence should not discourage the use of D-PSF as a predictor in kindergarten. Results require replication with other criterion measure of reading and analyses of predictive accuracy are needed to determine whether D-PSF in kindergarten may be predictive of reading in later grades despite the weak association with reading observed at the end of kindergarten. Measures of phonological awareness may be more strongly correlated with the criterion measures in second grade, when the complexity of reading content requires students who have relied on word-recognition strategies to use phonemic decoding strategies (Ehri & Soffer, 1999). The latter may be facilitated if students’ phonemic awareness skills are well-developed (Ehri, 1999). Within the outcomes-driven model, D-PSF is presumed to have instructional relevance: Scores predict future reading performance but also indicate goals for instruction related to phonemic awareness skills. Results from this study show that D-PSF
correlated to other DIBS strongest in winter and with weaker correlations in spring. The influence of phonemic awareness on other critical skills, particularly those involving letter-sound correspondences, may peak in the middle of kindergarten. Thus, phonemic awareness instruction in the first half of kindergarten may facilitate learning other necessary skills (e.g., letter-sound fluency, nonsense-word fluency) that demonstrate more direct relationships with reading outcomes.

Finally, predictive validity of A-LSF in kindergarten was demonstrated at all three benchmarks. In winter, explained considerable shared variance in criterion measures with D-NWF, and on its own, explained significant unique variance. The practical significance of these results in unclear. Comparisons of predictive accuracy between A-LSF and D-NWF may clarify whether A-LSF would constitute a meaningful addition to DIBELS or whether the efficiency of DIBELS would be compromised by a redundant measure. Perhaps most interesting was that scores on D-LNF were not explained by scores on A-LSF. This finding does not support the hypothesis that D-LNF is actually an indicator of letter-sound knowledge, and suggests that D-LNF measures knowledge or abilities that are relevant to early reading acquisition but that are distinct from letter-sound knowledge. In the spring, for example, D-LNF accounted for 9% of the unique variance in D-ORF. Considering that letter-naming skills has little instructional relevance (Adams, 1990), the strong relationship of D-LNF to reading outcomes suggests that D-LNF scores reflect performance on reading skills other than letter-sound fluency. Future research investigating the construct validity of D-LNF may reveal other critical early literacy skills that would function as indicators of reading, but with greater instructional relevance than D-LNF.

Limitations
Several limitations of the present study warrant consideration. Foremost, is that the results are based on correlational analyses. Some of the relationships observed among the measures are strong, and likely meaningful, but making causal inferences based on this evidence is unwise. Furthermore, the approach chosen to determine the optimal set of predictors, blockwise selection (Pedhazur, 1997) is not the optimal technique for doing so. The most effective empirical approach for identifying predictors is *all possible subsets regression* (Pedhazur, 1997), but the analyses and accurate interpretation of results are far more complex than blockwise selection. With the relatively few variables in this study, blockwise selection was likely sufficient. Other approaches to comparing and ranking predictors such as dominance analysis have been demonstrated successfully in other investigations (Schtschnieder, et al., 2004; Fuchs, et al., 2004) and may yield different results if applied to the data from the study.

Messick (1989) purported that evidence of predictive and concurrent validity is a necessary but insufficient foundation for establishing the construct validity of a measure. Clear evidence is necessary that the information obtained from a measure results in productive social outcomes. In the present study, examining the validity of DIBS in early literacy is well-intended; however, the study did not include evidence that demonstrates the utility of the measures. Multi-year studies are needed that examine the student outcomes that result from using DIBS in early literacy as a framework for educational decision-making.

The lack of sufficient reliability data also presents a limitation. Inter-rater reliability scores were collected at each benchmark, but in fall, only 22 sets of scores were collected and in spring, only 15 were collected. A more acceptable ratio is 20% (Salvia & Ysseldyke, 1995) of total observations, and in the present study this proportion is approximately 35 per benchmark.
Finally, no attempt was made to examine the effects of instruction that took place between benchmark assessments. Though growth between measurements was expected, the differences in the quality of instruction between classrooms and the nature of the school-wide curriculum at different points during the year was not considered in the analyses. Quality of instruction could be a potentially important factor influencing the relationship among the measures, particularly the relationships among measures at different benchmarks.

**Summary and Conclusions**

Identifying students in need of literacy intervention as early as possible is perhaps the most efficient and effective way to comply with new reform initiatives like the No Child Left Behind Act of 2000 that requires all students be competent readers by the third grade. Indeed, the prevention of reading failure promises the most meaningful outcomes for students who might otherwise suffer the life-long consequences of illiteracy. Dynamic indicators of basic skills, particularly DIBELS, developed by Good and Kaminski (2002), have received a great deal of attention from school systems seeking valid measures to use within preventative approaches to reading failure. The dynamic indicators of basic skills in early literacy investigated in this study were those from DIBELS and AimsWeb Letter-Sound Fluency (Shinn & Shinn, 2002). Together, these measures represent a range of component literacy skills that includes phonological awareness, letter-naming, letter-sound correspondence, sound blending, and fluency. All are hypothesized to be important to reading acquisition, and all have been used to predict reading later reading competence. The results of the predictive and concurrent analyses in the study suggest the general superiority of the print-related measures (D-NWF, D-LNF, & A-LSF) at all three kindergarten benchmarks in predicting reading outcomes. Results support D-LNF administered in the fall a strong predictor of reading. In winter and in spring of
kindergarten, D-NWF proved to be the strongest predictor. Also in winter, A-LSF made a slight but statistically significant contribution to explaining end of year reading outcomes. Whether this result justifies adding a measure of letter-sound fluency to the kindergarten DIBELS measures requires further examination. Despite the existing theoretical and empirical support, the predictive and concurrent validity of the two phonological awareness measures (D-PSF and D-ISF) was not strongly supported. Nevertheless, considering the preliminary nature of the results and the potential problems with reliability, inferences regarding the importance of phonological awareness within the context of measurement or instruction should be made with caution. The support that is demonstrated for the use of D-NWF and D-LNF in kindergarten is more encouraging. Future research should continue to explore the relative influences of component literacy skills, involving both speech and print, on eventual reading outcomes. A greater understanding of the multiple dimensions of early reading acquisition will ultimately lead to a set of valid instruments used to identify children at-risk of reading failure and to guide interventions that promote success.
REFERENCES


Burke, M. D., Hagan-Burke, S., & Ditkowsky, B. (in review). The predictive validity of initial sound fluency and letter naming fluency at the beginning of kindergarten with later reading outcomes.


deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.


APPENDIX A

DIBELS INITIAL SOUND FLUENCY
Appendix A

**Benchmark K-1**
DIBELS® Initial Sound Fluency

This is tomato, cub, plate, doughnut (point to pictures).

1. Which picture begins with /d/?
   0 1
2. Which picture begins with /t/?
   0 1
3. Which picture begins with /k/?
   0 1
4. What sound does “plate” begin with?
   0 1

This is bump, insect, refrigerator, skate (point to pictures).

5. Which picture begins with /sk/?
   0 1
6. Which picture begins with /r/?
   0 1
7. Which picture begins with /b/?
   0 1
8. What sound does “insect” begin with?
   0 1

This is rooster, mule, fly, soap (point to pictures).

9. Which picture begins with /r/?
   0 1
10. Which picture begins with /fl/?
    0 1
11. Which picture begins with /s/?
    0 1
12. What sound does “mule” begin with?
    0 1

This is pliers, doctor, quilt, beetle (point to pictures).

13. Which picture begins with /b/?
    0 1
14. Which picture begins with /pl/?
    0 1
15. Which picture begins with /d/?
    0 1
16. What sound does “quilt” begin with?
    0 1

Time: ___________ Seconds  Total Correct: ________

\[
\frac{60 \times \text{Total Correct}}{\text{Seconds}} = \text{Correct Initial Sounds per Minute}
\]
APPENDIX B

DIBELS PHONEME SEGMENTATION FLUENCY
### Benchmark K-2
DIBELS<sup>TM</sup> Phoneme Segmentation Fluency

<table>
<thead>
<tr>
<th>Word</th>
<th>Phonemes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>hat</td>
<td>/h/ /a/ /t/</td>
<td>6</td>
</tr>
<tr>
<td>hear</td>
<td>/h/ /ea/ /r/</td>
<td>6</td>
</tr>
<tr>
<td>as</td>
<td>/a/ /z/</td>
<td>6</td>
</tr>
<tr>
<td>punch</td>
<td>/p/ /u/ /n/ /ch/</td>
<td>6</td>
</tr>
<tr>
<td>means</td>
<td>/m/ /ea/ /n/ /z/</td>
<td>6</td>
</tr>
<tr>
<td>by</td>
<td>/b/ /ie/</td>
<td>6</td>
</tr>
<tr>
<td>seem</td>
<td>/s/ /ea/ /m/</td>
<td>6</td>
</tr>
<tr>
<td>ship</td>
<td>/sh/ /i/ /p/</td>
<td>6</td>
</tr>
<tr>
<td>ought</td>
<td>/o/ /t/</td>
<td>5</td>
</tr>
<tr>
<td>pack</td>
<td>/p/ /a/ /k/</td>
<td>5</td>
</tr>
<tr>
<td>jam</td>
<td>/j/ /a/ /m/</td>
<td>5</td>
</tr>
<tr>
<td>if</td>
<td>/i/ /f/</td>
<td>5</td>
</tr>
<tr>
<td>yell</td>
<td>/y/ /e/ /l/</td>
<td>6</td>
</tr>
<tr>
<td>ham</td>
<td>/h/ /a/ /m/</td>
<td>6</td>
</tr>
<tr>
<td>calls</td>
<td>/k/ /o/ /l/ /z/</td>
<td>6</td>
</tr>
<tr>
<td>ear</td>
<td>/ea/ /r/</td>
<td>6</td>
</tr>
<tr>
<td>key</td>
<td>/k/ /ea/</td>
<td>6</td>
</tr>
<tr>
<td>crowd</td>
<td>/k/ /r/ /ow/ /d/</td>
<td>6</td>
</tr>
<tr>
<td>loud</td>
<td>/l/ /ow/ /d/</td>
<td>6</td>
</tr>
<tr>
<td>choose</td>
<td>/ch/ /oo/ /z/</td>
<td>6</td>
</tr>
<tr>
<td>bare</td>
<td>/b/ /ai/ /r/</td>
<td>7</td>
</tr>
<tr>
<td>bills</td>
<td>/b/ /i/ /l/ /z/</td>
<td>7</td>
</tr>
<tr>
<td>guy</td>
<td>/g/ /ie/</td>
<td>7</td>
</tr>
<tr>
<td>stand</td>
<td>/s/ /t/ /a/ /n/ /d/</td>
<td>7</td>
</tr>
</tbody>
</table>

**Total:** 49

**Error Pattern:**
APPENDIX C

DIBELS LETTER-NAMING FLUENCY
Appendix C

Benchmark K-1
DIBELS™ Letter Naming Fluency

<table>
<thead>
<tr>
<th>g</th>
<th>N</th>
<th>E</th>
<th>Y</th>
<th>R</th>
<th>l</th>
<th>V</th>
<th>d</th>
<th>H</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>d</td>
<td>x</td>
<td>S</td>
<td>C</td>
<td>n</td>
<td>j</td>
<td>H</td>
<td>s</td>
<td>S</td>
</tr>
<tr>
<td>E</td>
<td>n</td>
<td>G</td>
<td>h</td>
<td>c</td>
<td>i</td>
<td>h</td>
<td>B</td>
<td>b</td>
<td>O</td>
</tr>
<tr>
<td>Y</td>
<td>F</td>
<td>p</td>
<td>D</td>
<td>L</td>
<td>i</td>
<td>q</td>
<td>c</td>
<td>D</td>
<td>Q</td>
</tr>
<tr>
<td>R</td>
<td>v</td>
<td>F</td>
<td>J</td>
<td>Z</td>
<td>M</td>
<td>P</td>
<td>o</td>
<td>p</td>
<td>u</td>
</tr>
<tr>
<td>l</td>
<td>G</td>
<td>A</td>
<td>f</td>
<td>V</td>
<td>B</td>
<td>P</td>
<td>k</td>
<td>m</td>
<td>I</td>
</tr>
<tr>
<td>V</td>
<td>M</td>
<td>e</td>
<td>r</td>
<td>y</td>
<td>z</td>
<td>a</td>
<td>L</td>
<td>U</td>
<td>A</td>
</tr>
<tr>
<td>d</td>
<td>y</td>
<td>q</td>
<td>v</td>
<td>w</td>
<td>u</td>
<td>T</td>
<td>w</td>
<td>N</td>
<td>U</td>
</tr>
<tr>
<td>H</td>
<td>j</td>
<td>K</td>
<td>e</td>
<td>r</td>
<td>X</td>
<td>T</td>
<td>z</td>
<td>Y</td>
<td>X</td>
</tr>
<tr>
<td>Z</td>
<td>x</td>
<td>f</td>
<td>m</td>
<td>W</td>
<td>W</td>
<td>s</td>
<td>J</td>
<td>I</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>E</td>
<td>R</td>
<td>K</td>
<td>g</td>
<td>N</td>
<td>E</td>
<td>Y</td>
<td>R</td>
<td>l</td>
</tr>
</tbody>
</table>

Total: _____
APPENDIX D

DIBELS NONSENSE WORD FLUENCY
Appendix D

Benchmark K-2
DIBELS™ Nonsense Word Fluency

y i z   w a n   z o c    f u l    m i k   ___/15
z u m  n u f    k u n   r u v    f o d   ___/15
v e p   i j   o p   j u j   s u g    ___/13
z u z   o v   v i t   w a m   b u k    ___/14
l e f   l u k   t e v   l o f   k o m    ___/15
j u f   t a m   n o l   r e z   k e c    ___/15
p u m   p o z   m u m   o l   k a v    ___/14
r i v   k i c   k i s   k e m   v a k    ___/15
t e k   u t   r i z   a j   v e j    ___/13
y i l   j e v   n e g   s o m   j u p    ___/15

Total: _____

Error Pattern:
APPENDIX E

AIMSWEB LETTER-SOUND FLUENCY
Appendix E

AIMSweb® Letter Sound Fluency - Benchmark Assessment #1 (Kindergarten - Fall)

Given To: __________________________  Given By: __________________________  Date: ____________

<table>
<thead>
<tr>
<th>t</th>
<th>d</th>
<th>n</th>
<th>r</th>
<th>p</th>
<th>c</th>
<th>z</th>
<th>v</th>
<th>w</th>
<th>k</th>
<th>/ 10 (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>b</td>
<td>t</td>
<td>f</td>
<td>v</td>
<td>z</td>
<td>i</td>
<td>c</td>
<td>d</td>
<td>p</td>
<td>/ 10 (20)</td>
</tr>
<tr>
<td>y</td>
<td>e</td>
<td>l</td>
<td>b</td>
<td>j</td>
<td>s</td>
<td>t</td>
<td>f</td>
<td>a</td>
<td></td>
<td>/ 10 (30)</td>
</tr>
<tr>
<td>c</td>
<td>n</td>
<td>f</td>
<td>r</td>
<td>m</td>
<td>b</td>
<td>t</td>
<td>h</td>
<td>z</td>
<td>s</td>
<td>/ 10 (40)</td>
</tr>
<tr>
<td>j</td>
<td>k</td>
<td>p</td>
<td>s</td>
<td>f</td>
<td>h</td>
<td>i</td>
<td>r</td>
<td>o</td>
<td>m</td>
<td>/ 10 (50)</td>
</tr>
<tr>
<td>s</td>
<td>z</td>
<td>p</td>
<td>i</td>
<td>j</td>
<td>r</td>
<td>e</td>
<td>d</td>
<td>g</td>
<td>o</td>
<td>/ 10 (60)</td>
</tr>
<tr>
<td>j</td>
<td>g</td>
<td>a</td>
<td>t</td>
<td>s</td>
<td>h</td>
<td>c</td>
<td>r</td>
<td>k</td>
<td>l</td>
<td>/ 10 (70)</td>
</tr>
<tr>
<td>j</td>
<td>u</td>
<td>k</td>
<td>y</td>
<td>a</td>
<td>s</td>
<td>z</td>
<td>e</td>
<td>i</td>
<td>v</td>
<td>/ 10 (80)</td>
</tr>
<tr>
<td>m</td>
<td>s</td>
<td>d</td>
<td>g</td>
<td>f</td>
<td>l</td>
<td>b</td>
<td>v</td>
<td>j</td>
<td>c</td>
<td>/ 10 (90)</td>
</tr>
<tr>
<td>t</td>
<td>e</td>
<td>m</td>
<td>l</td>
<td>w</td>
<td>j</td>
<td>y</td>
<td>z</td>
<td>f</td>
<td>v</td>
<td>/ 10 (100)</td>
</tr>
</tbody>
</table>
APPENDIX F

DIBELS ORAL READING FLUENCY
Appendix F

Benchmark 2.1
DIBELS™ Oral Reading Fluency

Spring is Coming

It has been so cold this winter. The wind blew and blew. It rained and rained. The days have been gray and dark. I had to wear mittens and a hat to school every day. It even snowed twice.

At first winter was fun. Now I’m tired of the cold. It has been too cold and wet to play outside. At school, we sit in the library and read during recess. After school I just stay in the house and play. I don’t want to play inside anymore.

But today was nice. The sun was shining brightly even though it was still cold. The wind didn’t blow. My friends and I played kick ball at recess. We had to take off our jackets because we were warm. We even got hot and thirsty.

On the way home from school I saw a purple flower on our street. It was blooming in the grass. I told my mother about it. She wanted me to show it to her. She bent down and touched it.

“Come sniff this,” she said. It smelled like perfume and sun all mixed together. “Spring must be right around the corner,” she said. “This is a crocus. It’s one of the first flowers of spring.”

I can’t wait for spring.

Retell: Total:
<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Reliability Evidence</th>
<th>Measure</th>
<th>Stability</th>
<th>Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliot, Lee, &amp; Tollefson (2001).</td>
<td>Test-retest, two-day interim. Equivalent-forms, two-day interim</td>
<td>D-LNF</td>
<td>.90</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A-LSF</td>
<td>.83</td>
<td>.82</td>
</tr>
<tr>
<td>Tindal, Marston, &amp; Deno (1983).</td>
<td>Test-retest, two week interim. Equivalent-forms, two week interim.</td>
<td>CBM-ORF</td>
<td>.92 to .97</td>
<td>.84 to .94</td>
</tr>
<tr>
<td>Good &amp; Kaminski (1996).</td>
<td>Equivalent forms, one-month interim.</td>
<td>D-ISF</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D-LNF</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D-NWF</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>Good, Simmons, &amp; Kameneui (2001)</td>
<td></td>
<td>D-ISF</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D-PSF</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D-LNF</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Good, Kaminski, Laimon, &amp; Johnson (2000); Kaminski &amp; Good (1996). Equivalent-forms</td>
<td>D-LNF</td>
<td>.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D-PSF</td>
<td>.88</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Letter Sound Fluency (A-LSF) is not a standard DIBELS measure.
<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Validity Evidence</th>
<th>Criterion Measure</th>
<th>Measure</th>
<th>Coefficients Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good, Simmons, &amp; Kame’enui (2001).</td>
<td>Predictive.</td>
<td>Woodcock-Johnson Revised Reading Readiness Cluster scores (1st grade).</td>
<td>D-ISF (K)</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-PSF(K)</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-LNF(K)</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-NWF(K)</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIBELS subtests at later benchmarks:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. D-PSF (spring K)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. D-NWF (winter 1st)</td>
<td>a. D-ISF (winter K)</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. D-ORF (spring 1st)</td>
<td>b. D-PSF (spring K)</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. D-NWF (winter 1st)</td>
<td>.78</td>
</tr>
<tr>
<td>Good, Kaminski, Laimon, &amp; Johnson (2000); Kaminski &amp; Good (1996).</td>
<td>Concurrent criterion-related</td>
<td>Metropolitan Readiness Test</td>
<td>D-LNF(K)</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-PSF(K)</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McCarthy Scales of Children’s Abilities</td>
<td>D-LNF(K)</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td>Predictive</td>
<td>Stanford Diagnostic Reading Test</td>
<td>D-LNF(K)</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CBM-ORF</td>
<td>D-LNF(K)</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-PSF(K)</td>
<td>.76</td>
</tr>
<tr>
<td>Author</td>
<td>Type of Validity Evidence</td>
<td>Criterion Measure</td>
<td>Measure</td>
<td>Coefficients Reported</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Elliot, Lee, &amp; Toleffson (2001).</td>
<td>Test of Phonological Awareness</td>
<td>D-LNF</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A-LSF</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kaufman Brief Intelligence Test, Verbal</td>
<td>D-LNF</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A-LSF</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>Hagan-Burke, Burke, &amp; Crowder (in press)</td>
<td>Concurrent criterion-related</td>
<td>Test of Word Reading Efficiency, (TOWRE): Sight Words and Phoneme Decoding Efficiency subtests</td>
<td>D-PSF</td>
<td>.32 (SW) .29 (PDE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-LNF</td>
<td>.62 (SW) .46 (PDE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-NWF</td>
<td>.73 (SW) .74 (PDE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WUF</td>
<td>.32 (SW) .29 (PDE)</td>
</tr>
<tr>
<td></td>
<td>Predictive</td>
<td>Fall DIBELS subtest scores predicting TOWRE scores in Winter</td>
<td>D-NWF</td>
<td>D-NWF and D-LNF accounted for the majority of variance in both TOWRE subtest scores.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-LNF</td>
<td></td>
</tr>
<tr>
<td>VanDerHeyden, Witt, Naquin, &amp; Noell (2001)</td>
<td>Concurrent criterion-related</td>
<td>Experimental outcome measures of basil skills-kindergarten: Reading-Circle Letter</td>
<td>D-ISF</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Letter Sound Naming</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Math-Circle Number</td>
<td>D-ISF</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Letter Sound Naming</td>
<td>.70</td>
</tr>
</tbody>
</table>

*Note. Letter Sound Fluency (A-LSF) and Letter Sound Naming are not standard DIBELS measures.*
<table>
<thead>
<tr>
<th>Measure</th>
<th>FALL SCORES</th>
<th>Winter Scores</th>
<th>Spring Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk Category</td>
<td>Risk Category</td>
<td>Risk Category</td>
</tr>
<tr>
<td>D-ISF</td>
<td>At-Risk &lt;4</td>
<td>Some-Risk 4-7</td>
<td>Low-Risk ≥8</td>
</tr>
<tr>
<td></td>
<td>At-Risk &lt;10</td>
<td>Some-Risk 10-24</td>
<td>Low-Risk ≥25</td>
</tr>
<tr>
<td>D-LNF</td>
<td>Some-Risk &lt;2</td>
<td>Some-Risk 2-7</td>
<td>Some-Risk ≥8</td>
</tr>
<tr>
<td></td>
<td>At-Risk &lt;15</td>
<td>At-Risk 15-26</td>
<td>At-Risk ≥27</td>
</tr>
<tr>
<td>D-PSF</td>
<td>At-Risk &lt;7</td>
<td>At-Risk 1-17</td>
<td>Some-Risk ≥18</td>
</tr>
<tr>
<td></td>
<td>At-Risk &lt;10</td>
<td>At-Risk 10-34</td>
<td>Low-Risk &gt;35</td>
</tr>
<tr>
<td>D-NWF</td>
<td>At-Risk &lt;5</td>
<td>At-Risk 5-12</td>
<td>Some-Risk ≥13</td>
</tr>
<tr>
<td></td>
<td>At-Risk &lt;15</td>
<td>Some-Risk 15-24</td>
<td>Low-Risk &gt;25</td>
</tr>
</tbody>
</table>

Norm-Group Mean Scores for A-LSF Kindergarten

<table>
<thead>
<tr>
<th></th>
<th>FALL</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample M</td>
<td>Norm M</td>
<td>Sample M</td>
<td>Norm M</td>
</tr>
<tr>
<td>A-LSF</td>
<td>4.42</td>
<td>7</td>
<td>20.95</td>
</tr>
</tbody>
</table>
Table 4.  
*Participant Characteristics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>89</td>
<td>49.7</td>
</tr>
<tr>
<td>Female</td>
<td>88</td>
<td>50.3</td>
</tr>
<tr>
<td>Race / Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>60</td>
<td>33.9</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>Caucasian</td>
<td>107</td>
<td>60.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>Multi-Racial</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistance</td>
<td>66</td>
<td>37.2</td>
</tr>
<tr>
<td>Partial assistance</td>
<td>5</td>
<td>2.8</td>
</tr>
<tr>
<td>No assistance</td>
<td>106</td>
<td>59.8</td>
</tr>
<tr>
<td>Special Education services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receives services</td>
<td>27</td>
<td>15.3</td>
</tr>
<tr>
<td>No special services</td>
<td>150</td>
<td>84.7</td>
</tr>
</tbody>
</table>

*Note.* Socio-economic status is indicated by level of financial assistance with school lunch.
<table>
<thead>
<tr>
<th></th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DIBELS Subtests Indicated for</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Kindergarten</em></td>
<td>D-ISF</td>
<td>D-ISF</td>
<td>D-LNF</td>
</tr>
<tr>
<td></td>
<td>D-LNF</td>
<td>D-LNF</td>
<td>D-PSF</td>
</tr>
<tr>
<td></td>
<td>D-PSF</td>
<td>D-NWF</td>
<td>D-NWF</td>
</tr>
<tr>
<td><strong>Letter-Sound Fluency</strong></td>
<td>A-LSF</td>
<td>A-LSF</td>
<td>A-LSF</td>
</tr>
<tr>
<td><strong>Criterion Measures of Reading Ability</strong></td>
<td>None</td>
<td>None</td>
<td>TERA-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DIBELS-ORF</td>
</tr>
</tbody>
</table>
Table 6.
Reliability Estimates for DIBS and Spring Criterion Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Test-Retest Reliability—Fall</th>
<th>Interrater Reliability—Fall</th>
<th>Interrater Reliability—Winter</th>
<th>Interrater Reliability—Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N=50)</td>
<td>(N=22)</td>
<td>(N=35)</td>
<td>(N=15)</td>
</tr>
<tr>
<td>Initial Sound Fluency</td>
<td>.72</td>
<td>.85</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Letter-Naming Fluency</td>
<td>.91</td>
<td>.98</td>
<td>.95</td>
<td>.99</td>
</tr>
<tr>
<td>Letter-Sound Fluency</td>
<td>.82</td>
<td>.88</td>
<td>.90</td>
<td>.89</td>
</tr>
<tr>
<td>Phoneme Segmentation Fluency</td>
<td></td>
<td></td>
<td>.91</td>
<td>.93</td>
</tr>
<tr>
<td>Nonsense Word Fluency</td>
<td></td>
<td></td>
<td>.95</td>
<td>.94</td>
</tr>
</tbody>
</table>
Table 7. *Descriptive Statistics*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SEM</th>
<th>SD</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISF1</td>
<td>0</td>
<td>30</td>
<td>9.91</td>
<td>.482</td>
<td>6.417</td>
<td>41.174</td>
<td>.835</td>
<td>.643</td>
</tr>
<tr>
<td>LNF1</td>
<td>0</td>
<td>54</td>
<td>14.02</td>
<td>.979</td>
<td>13.019</td>
<td>169.488</td>
<td>.788</td>
<td>-.195</td>
</tr>
<tr>
<td>LSF1</td>
<td>0</td>
<td>25</td>
<td>4.42</td>
<td>.447</td>
<td>5.947</td>
<td>35.371</td>
<td>1.533</td>
<td>1.577</td>
</tr>
<tr>
<td>ISF2</td>
<td>2</td>
<td>50</td>
<td>18.12</td>
<td>.696</td>
<td>9.255</td>
<td>85.662</td>
<td>.772</td>
<td>.558</td>
</tr>
<tr>
<td>LNF2</td>
<td>1</td>
<td>82</td>
<td>34.30</td>
<td>1.139</td>
<td>15.147</td>
<td>229.438</td>
<td>.035</td>
<td>-.007</td>
</tr>
<tr>
<td>PSF2</td>
<td>0</td>
<td>57</td>
<td>16.65</td>
<td>1.181</td>
<td>15.716</td>
<td>247.002</td>
<td>.791</td>
<td>-.455</td>
</tr>
<tr>
<td>NWF2</td>
<td>0</td>
<td>72</td>
<td>17.75</td>
<td>1.059</td>
<td>14.084</td>
<td>198.350</td>
<td>.759</td>
<td>.755</td>
</tr>
<tr>
<td>LSF2</td>
<td>0</td>
<td>63</td>
<td>20.95</td>
<td>1.010</td>
<td>13.432</td>
<td>180.424</td>
<td>.518</td>
<td>.075</td>
</tr>
<tr>
<td>LNF3</td>
<td>6</td>
<td>86</td>
<td>42.99</td>
<td>1.066</td>
<td>14.184</td>
<td>201.182</td>
<td>.256</td>
<td>.039</td>
</tr>
<tr>
<td>PSF3</td>
<td>0</td>
<td>66</td>
<td>32.88</td>
<td>1.054</td>
<td>14.020</td>
<td>196.564</td>
<td>-.266</td>
<td>-.235</td>
</tr>
<tr>
<td>NWF3</td>
<td>0</td>
<td>78</td>
<td>28.62</td>
<td>1.143</td>
<td>15.212</td>
<td>231.418</td>
<td>.874</td>
<td>.576</td>
</tr>
<tr>
<td>LSF3</td>
<td>0</td>
<td>65</td>
<td>29.90</td>
<td>.924</td>
<td>12.297</td>
<td>151.206</td>
<td>.306</td>
<td>-.369</td>
</tr>
<tr>
<td>ORF3</td>
<td>0</td>
<td>34</td>
<td>7.50</td>
<td>.609</td>
<td>8.099</td>
<td>65.592</td>
<td>1.170</td>
<td>.767</td>
</tr>
<tr>
<td>TERA-3</td>
<td>22.00</td>
<td>53.00</td>
<td>39.1073</td>
<td>.43934</td>
<td>5.84504</td>
<td>34.165</td>
<td>-.136</td>
<td>-.209</td>
</tr>
</tbody>
</table>

*Note.* Numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring.
Table 8.  
*Correlation Matrix of DIBS and Criterion Measures of Reading*

<table>
<thead>
<tr>
<th></th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ISF1</td>
<td>LNF1</td>
<td>A-LSF1</td>
</tr>
<tr>
<td>ISF1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNF1</td>
<td>.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISF2</td>
<td></td>
<td>.27</td>
<td>.62</td>
</tr>
<tr>
<td>LNF2</td>
<td>.33</td>
<td>.61</td>
<td>.73</td>
</tr>
<tr>
<td>PSF2</td>
<td>.34</td>
<td>.40</td>
<td>.71</td>
</tr>
<tr>
<td>NWF2</td>
<td>.31</td>
<td>.48</td>
<td>.41</td>
</tr>
<tr>
<td>A-LSF2</td>
<td>.24</td>
<td>.48</td>
<td>.41</td>
</tr>
<tr>
<td>LNF3</td>
<td>.24</td>
<td>.36</td>
<td>.43</td>
</tr>
<tr>
<td>PSF3</td>
<td>.15*</td>
<td>.35</td>
<td>.38</td>
</tr>
<tr>
<td>NWF3</td>
<td>.24</td>
<td>.30</td>
<td>.32</td>
</tr>
<tr>
<td>A-LSF3</td>
<td>.26</td>
<td>.26</td>
<td>.27</td>
</tr>
<tr>
<td>D-ORF</td>
<td>.26</td>
<td>.27</td>
<td>.29</td>
</tr>
<tr>
<td>TERA-3</td>
<td>.15*</td>
<td>.55</td>
<td>.58</td>
</tr>
</tbody>
</table>
| Note: All correlations are significant at the 0.01 level except the TERA-3 and ISF1, which is significant at the .05 level. The numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring.
Table 9.  
*Correlations Among DIBS from Fall, Winter, and Spring Benchmarks and D-ORF with TERA-3 Subtests*

<table>
<thead>
<tr>
<th>TERA-3 Subtests</th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ISF1</td>
<td>LNF1</td>
<td>LSF1</td>
</tr>
<tr>
<td>TERA-Alphabet</td>
<td>.16*</td>
<td>.38**</td>
<td>.32**</td>
</tr>
<tr>
<td>TERA-Conventions</td>
<td>.03</td>
<td>.35**</td>
<td>.29**</td>
</tr>
<tr>
<td>TERA-Meaning</td>
<td>.18*</td>
<td>.27**</td>
<td>.36**</td>
</tr>
</tbody>
</table>

*Note.* Numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring. **p < .01, * p < .05.
Table 10. Predicting D-ORF and TERA-3 from Fall DIBS: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression

<table>
<thead>
<tr>
<th>Fall Predictors</th>
<th>D-ORF (Spring)</th>
<th>TERA-3 (Spring)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>R² Change</td>
</tr>
<tr>
<td>Bivariate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-ISF1</td>
<td>.06</td>
<td>--</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-LNF1</td>
<td>.29</td>
<td>--</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-LSF1</td>
<td>.19</td>
<td>--</td>
</tr>
<tr>
<td>Hierarchical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. D-ISF1</td>
<td>.06</td>
<td>--</td>
</tr>
<tr>
<td>b. A-LSF1</td>
<td>.30</td>
<td>.24</td>
</tr>
<tr>
<td>c. D-LNF1</td>
<td>.31</td>
<td>.01</td>
</tr>
<tr>
<td>Forward Blockwise Selection:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 5: (D-ORF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-LNF1 (D-ISF1, A-LSF1 removed)</td>
<td>.29</td>
<td>--</td>
</tr>
<tr>
<td>Model 6: (TERA-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-LNF1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-LSF1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reversed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-LSF1</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>D-LNF1</td>
<td>.21</td>
<td>.05</td>
</tr>
</tbody>
</table>

*Note.* Numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring. **p < .01, *p < .05.
Table 11

Predicting D-ORF and TERA-3 from Winter DIBELS and A-LSF: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression

<table>
<thead>
<tr>
<th>Winter Predictors</th>
<th>Criterion Measures</th>
<th>D-ORF</th>
<th></th>
<th></th>
<th>TERA-3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bivariate</td>
<td>R²</td>
<td>R² Change</td>
<td>F Change</td>
<td>Beta</td>
<td>T</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-ISF2</td>
<td>Model 2</td>
<td>.08</td>
<td>--</td>
<td>16.603**</td>
<td>.294</td>
<td>4.075**</td>
</tr>
<tr>
<td>D-PSF2</td>
<td>Model 3</td>
<td>.18</td>
<td>--</td>
<td>39.910</td>
<td>.431</td>
<td>6.317**</td>
</tr>
<tr>
<td>D-NWF2</td>
<td>Model 4</td>
<td>.48</td>
<td>--</td>
<td>163.974**</td>
<td>.696</td>
<td>12.085**</td>
</tr>
<tr>
<td>D-LNF2</td>
<td>Model 5</td>
<td>.33</td>
<td>--</td>
<td>89.246**</td>
<td>.581</td>
<td>9.447**</td>
</tr>
<tr>
<td>A-LSF2</td>
<td>Model 6</td>
<td>.42</td>
<td>--</td>
<td>132.216**</td>
<td>.656</td>
<td>11.499**</td>
</tr>
<tr>
<td>Hierarchical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. D-ISF2</td>
<td></td>
<td>.08</td>
<td>.08</td>
<td>16.603**</td>
<td>-.035</td>
<td>-.576</td>
</tr>
<tr>
<td>b. D-PSF2</td>
<td></td>
<td>.19</td>
<td>.11</td>
<td>25.417**</td>
<td>.023</td>
<td>.332</td>
</tr>
<tr>
<td>c. A-LSF2</td>
<td></td>
<td>.43</td>
<td>.24</td>
<td>75.645**</td>
<td>.220</td>
<td>2.251**</td>
</tr>
<tr>
<td>d. D-NWF2</td>
<td></td>
<td>.49</td>
<td>.06</td>
<td>22.605**</td>
<td>.441</td>
<td>4.202**</td>
</tr>
<tr>
<td>e. D-LNF2</td>
<td></td>
<td>.50</td>
<td>.00</td>
<td>1.856</td>
<td>.111</td>
<td>1.362</td>
</tr>
</tbody>
</table>
Table 11 Continued

<table>
<thead>
<tr>
<th>Forward Blockwise Selection</th>
<th>R²</th>
<th>R² Change</th>
<th>F Change</th>
<th>Beta</th>
<th>T</th>
<th>R²</th>
<th>R² Change</th>
<th>F Change</th>
<th>Beta</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 8 (D-ORF)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF2</td>
<td>.48</td>
<td>--</td>
<td>163.974**</td>
<td>.477</td>
<td>5.272**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-LSF2</td>
<td>.50</td>
<td>.02</td>
<td>8.888**</td>
<td>.270</td>
<td>2.981**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reversed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-LSF2</td>
<td>.42</td>
<td></td>
<td>132.216**</td>
<td>.270</td>
<td>2.981**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF2</td>
<td>.50</td>
<td>.07</td>
<td>27.794**</td>
<td>.477</td>
<td>5.272**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 9: TERA-3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF2</td>
<td>.35</td>
<td></td>
<td>98.823**</td>
<td>.314</td>
<td>3.335**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-LNF2</td>
<td>.38</td>
<td>.03</td>
<td>9.227**</td>
<td>.258</td>
<td>3.127**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-PSF2</td>
<td>.40</td>
<td>.02</td>
<td>5.518*</td>
<td>.172</td>
<td>2.349*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reversed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF2</td>
<td>.35</td>
<td></td>
<td>98.823**</td>
<td>.314</td>
<td>3.335**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-PSF2</td>
<td>.37</td>
<td>.02</td>
<td>4.953*</td>
<td>.258</td>
<td>3.127**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-LNF2</td>
<td>.40</td>
<td>.03</td>
<td>0.781**</td>
<td>.172</td>
<td>2.349*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring. ** p < .01, * p < .05.*
### Table 12

**Concurrent Validity of Spring DIBELS and A-LSF with D-ORF and TERA-3: Model Summaries for Bivariate, Hierarchical, and Forward Blockwise Regression**

<table>
<thead>
<tr>
<th>Spring Predictors</th>
<th>D-ORF</th>
<th>Criterion Measures</th>
<th>TERA-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bivariate</td>
<td>D-PSF3</td>
<td>R^2</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td>D-PSF3</td>
<td>.09</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td>D-NWF3</td>
<td>.43</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td>D-LNF3</td>
<td>.40</td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
<td>A-LSF3</td>
<td>.27</td>
</tr>
<tr>
<td>Hierarchical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td></td>
<td>a. D-PSF3</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. A-LSF3</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. D-NWF3</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. D-LNF3</td>
<td>.52</td>
</tr>
</tbody>
</table>
Table 12 Continued

<table>
<thead>
<tr>
<th>Blockwise Selection</th>
<th>R²</th>
<th>R² Change</th>
<th>F Change</th>
<th>Beta</th>
<th>T</th>
<th>R²</th>
<th>R² Change</th>
<th>F Change</th>
<th>Beta</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 6 (D-ORF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF3</td>
<td>.43</td>
<td></td>
<td>135.396</td>
<td>.439</td>
<td>6.913**</td>
<td>.52</td>
<td></td>
<td>36.126</td>
<td>.382</td>
<td>6.011**</td>
</tr>
<tr>
<td>D-LNF3 (reversed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
<td></td>
<td>119.070**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.52</td>
<td></td>
<td>47.792**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 7 (TERA-3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-NWF3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.26</td>
<td></td>
<td>61.936**</td>
<td>.341</td>
<td>4.444**</td>
</tr>
<tr>
<td>D-LNF3 (reversed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.31</td>
<td>.05</td>
<td>14.658**</td>
<td>.294</td>
<td>3.829**</td>
</tr>
<tr>
<td>D-LNF3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.23</td>
<td>.24</td>
<td>55.711**</td>
<td>.341</td>
<td>4.444**</td>
</tr>
<tr>
<td>D-NWF3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.31</td>
<td>.07</td>
<td>19.748**</td>
<td>.294</td>
<td>3.829**</td>
</tr>
</tbody>
</table>

*Note.* Numbers next to variable names indicate season of administration: 1 = fall, 2 = winter, 3 = spring. ** p < .01, * p < .0