1-1-2003

The Utility of DIBELS as a Curriculum Based Measurement in Relation to Reading Proficiency on High Stakes Tests

Rebecca Gayle Cook

The Utility of DIBELS as a Curriculum Based Measurement in Relation to Reading Proficiency on High Stakes Tests

Thesis Submitted in Fulfillment

of the Requirements for the

Degree of Master of Arts in Psychology

Rebecca Gayle Cook

Marshall University Graduate College

August 25, 2003

Abstract

Rebecca Gayle Cook

The purpose of the current study was to investigate the relationship between the Dynamic Indicators of Basic Early Literacy Skills or DIBELS reading measures and the Stanford Achievement Test, Ninth Edition or SAT9 reading scores. The following research question was examined: What is the concurrent validity of the DIBELS reading measures in relation to students' SAT9 reading scores? In this study, archival data from five first-grade classrooms at a rural southeastern Ohio elementary school were collected. The results of the study indicated that there was a positive and significant correlation between DIBELS measures and SAT9 scores with the exception of the DIBELS Phoneme Segmentation Fluency or PSF reading measure and the SAT9 Word Reading subtest. The DIBELS Oral Reading Fluency or ORF measure is the subtest with the highest concurrent validity in relation to SAT9 reading scores. Conclusions and recommendations for further research were discussed.

Dedication

This research paper is dedicated to my husband, family, and friends for their solid support and encouragement throughout my journey of higher education. They have made it possible for me to achieve this pivotal moment in my life. When I procrastinated, they were there to push me forward. When I accomplished each milestone, they were my biggest cheerleaders. When I felt discouraged, they reminded me of my past accomplishments. In addition, I would also like to dedicate this accomplishment in memory of my grandfather who always emphasized the importance of striving for higher education. Thank you, one and all, for always being there when I needed you the most.

Acknowledgements

I would like to express my deep gratitude to my thesis committee, Dr. Elizabeth Kelley Boyles, Dr. Debra Lilly, and Dr. Fred Jay. Krieg. Thank you—for your guidance and constant support. A special word of thanks goes to Dr. Lilly for all of her statistical expertise, advice, and encouragement—every step of the way.

Table of Contents

List of Figures

The Utility of DIBELS as a Curriculum Based Measurement in Relation to Reading Proficiency on High Stakes Tests

Introduction to Review of Literature

With the preponderance of high stakes testing—or educational accountability with rewards and sanctions determined through standards-based assessment—and the provisions of the Early and Secondary Education Act (ESEA) signed into law on January 8, 2002, never before have educators been held so accountable for their students' academic achievement (Amrein & Berliner, 2003; US Dept of Education, 2003). Furthermore, President Bush has declared that all children will have achieved grade level reading skills by the end of their third grade year (Good, Simmons, & Kame'enui, 2001; Christie, 2001; US Dept of Education, 2003). With such a lofty goal, educators are now scrambling to find the "miracle cure" or the most effective research-based reading assessments and programs that will ensure their students achieve reading success.

The following paragraphs provide an overview of the importance of literacy, the latest instructional trends, and a discussion of two very different types of reading assessments and explain their utility in the classroom. Due to the amount of space and time, this review of the literature is not by any means exhaustive on the topic of reading. However, it is an attempt to present to the reader the reasoning behind the current study.

Importance of Literacy

Educators have long been aware that learning to read is a very involved, often times difficult, skill for some children to grasp. Lyon suggests for up to 20-30% of America's children, learning to read is one of the most difficult tasks that they will have to master in their life (1997). Considering that reading is a skill necessary for a child's

success both in school and in life, there is no question as to why low reading achievement is correlated with various social problems such as high school dropout rates, teen pregnancy, delinquency, unemployment, and homelessness (Kaminski & Good, 1996). In addition, poor reading skills have been linked to the development of behavioral and emotional problems including aggressive behavior, hyperactive behavior, poor self-concept, and a sense of hopelessness (Good & Simmons, 1998).

To further explore the repercussions of illiteracy, one would not have to search very hard or very long in the available literature. For example, according to the National Adult Literacy Survey, 40 million adults in the United States have low literacy skills and struggle with reading and helping their children with homework (International Literacy Network [ILN], 2003). Furthermore, the International Literacy Network indicates that literacy is the "ultimate gateway out of poverty" (ILN, 2003). For example, in the United States, workers without a diploma reportedly earn three times less income than those with a bachelor's degree. In addition, eight out of 20 Americans with low literacy skills live in poverty as compared to 1 in 20 Americans with strong literacy skills (National Institute for Literacy, 2003). Cumulative research also indicates that a child's literacy levels and motivation to stay in school is influenced by their parents' educational achievement. Children whose parents are unemployed and have not completed high school are five times more likely to drop out of school than children of working parents (ILN, 2003). Illiteracy also has its hidden costs. Low literate adults tend to be less healthy because they lack information of where to go, when to seek help, and are unable to read important information such as traffic signs, prescription information, and directions on baby formula. Over 70% of America's prisoners have low literacy skills and cannot perform

basic reading and writing tasks such as writing a letter or understanding a bus schedule (ILN, 2003; Stollar, 2002). Given these grim statistics, and the common knowledge that reading is the gateway to further learning, the ability to explore and learn in one's world independently, it is no wonder that reading achievement has been an important topic of research among educators for some time.

Effective Reading Instruction

The accumulation of years of such research can be found in a published work by the Partnership for Reading. This book, *Put Reading First, The Research Building Blocks for Teaching Children to Read*, summarizes what researchers have discovered to successfully teach children to read. It describes the findings of the National Reading Panel 2000 report on the five crucial areas of reading instruction: phonemic awareness, phonics or alphabetic principle, fluency, vocabulary, and text comprehension—also known as the Big Ideas in Beginning Reading (Big Ideas, 2003; Partnership for Reading, 2001; Stollar, 2002).

Phonemic awareness is the ability to hear and manipulate sounds in words. Alphabetic principle is the ability to associate sounds with letters and use the sounds to form words. Fluency is the effortless automatic ability to read words in connected text. Vocabulary is the ability to understand (receptive) and use (expressive) words to acquire and convey meaning. Finally, comprehension is the complex cognitive process involving the interaction between reading and text to convey meaning. In short, Big Ideas are pre-literacy skills and strategies that are the prerequisite and fundamental building blocks to later reading success. These crucial skills differentiate successful from less successful

readers and most importantly are found to be subject to change through instruction (Good, Simmons, & Kame'enui, 2001).

Unfortunately, research tells us that remediation of these skills is largely unsuccessful. In fact, a few studies mentioned by Kaminski and Good (1996) discuss the persistence of reading problems over time. In particular, one study investigated the reading and writing development of 54 children as they progressed from first through fourth grade. In that study, they found that the probability of a poor reader in the first grade remaining a poor reader in the fourth grade was .88 (Kaminski & Good, 1996). Furthermore, similar studies documented little improvement in problem readers between second and fifth grade. In yet another study, or review of existing research on the remediation of reading difficulties, it was found that even when remedial services (including Chapter I and Special Education) are provided, they are not very effective (1996). One would expect then, by third or fourth grade, for those students who are performing well below their peers, it would be too late to modify beginning reading instruction in order to promote the acquisition of initial reading skills (Good, Gruba & Kaminski, 2002). Therefore, the most sensible way to improve reading is to prevent reading problems from occurring in the first place. For instance, recent studies have shown when students with severe reading problems are given early, intensive instruction, nearly 95% can reach the national average in reading ability (Council for Exceptional Children, 1997). In addition to this, Lyon (1997), in summarizing 15 years of research for the National Institute of Child Health and Human Development reported:

> We have learned that for 85 to 90 percent of poor readers, prevention and early intervention programs that combine instruction in phoneme awareness, phonics, spelling, reading fluency, and reading comprehension strategies provided by well-trained teachers can increase reading skills to average reading levels. (p.1)

Given this, the focus on reading instruction among educators has become a diagnostic style of teaching. Early detection of poor readers through assessment, followed by intense instruction around the Big Ideas, are paramount to future reading success.

Stanford Achievement Test-Ninth Edition

High stakes testing or what some states call proficiency tests or group administered achievement tests is considered by some to be one way to track student achievement. However, these tests are often tied to some very serious consequences. For example, high stakes tests are widely used in 28 states as a means to determine grade promotion or high school graduation (Education USA, 2003).

One such group administered achievement test is the Stanford Achievement Test-Ninth Edition. The Stanford Achievement Test-Ninth Edition (SAT9) is a norm referenced, group administered achievement test for grades K.0 to 13.0. This test offers measures in reading, language, spelling, study skills, listening, mathematics, science and social science. The SAT9's standardizations were based on stratified random samples of 250,000 students from 1,000 school districts during the spring of 1995, and another 200,000 students during the fall. The stratification variables were socioeconomic status, urbanicity, and ethnicity. Students attending Catholic and other private schools were also

included. A total of 49 states and the District of Columbia were represented in the standardizations (Berk, 2003).

According to Berk, the SAT9's content was derived from an analyses of the most recent editions of textbooks in the relevant subject areas; the most recent state and district school curricula and objectives; and the most important trends in education according to such national professional organizations as the National Assessment of Educational Progress, Curriculum and Evaluation Standards for School Mathematics, Writing Process Model, National Science Education Standards, and National Council for the Social Studies Curriculum Standards (2003). Even with these efforts to reflect national core standards, Berk cautions school districts to judge the content against their own standards, stating that this criterion should be a primary consideration when reviewing the SAT9 (2003).

The SAT9's reliability—or the degree to which test scores are consistent, dependable, and repeatable—Kuder-Richardson Formula 20 (K-R20) coefficients are considered to be in the acceptable range (.80's to .90's for most multiple choice tests and .70's to low .80's in Listening, Language, Science and Social Science subtests) for making individual decisions about students (Berk, 2003; Haladyna, 2003). The validity of the SAT9, or the degree to which a certain inference from a test is appropriate or meaningful, was examined in two ways: content and construct. Content validity is bias or stereotyping in terms of gender, ethnicity, culture, socioeconomic status, and geographic region. According to Berk, the entire battery was brought before an advisory panel of minority-group educators to identify, revise or eliminate certain items (2003). In addition to this, comprehensive quantitative analyses using the Mantel-Haenszel statistic were

conducted for gender and Caucasian, African-American, and Hispanic student sample comparisons. The questionable items were then either revised or excluded, assuring that all items are valid for all examinees (Berk, 2003). Berk also reports on the construct validity as correlations between the SAT9 multiple-choice subtests and the Otis-Lennon School Ability Test. This correlation demonstrates the interrelationship between school achievement and ability. However, correlations between the SAT9 and other achievement batteries such as the California Achievement Tests and Iowa Tests of Basic Skills were not included and would have been more informative for this purpose (2003).

The items important for this study, the SAT9 reading clusters, Word Study, Word Reading, Reading Comprehension and Total Reading were explored. The Word Study Cluster contains three subtests namely Structural Analysis, Phonetic Analysis-Consonants, and Phonetic Analysis-Vowels offering 36 multiple-choice questions. Word Reading offers 30 multiple-choice questions. The Reading Comprehension cluster contains two-sentence stories, and short passages (cloze and questions). The short passages encompass recreational, textual, functional, initial understanding, and interpretation multiple-choice questions totaling 40. Finally, the Total Reading score is comprised of the three mentioned above cluster scores (Harcourt Educational Measurement, 2003). It would be fair to say that the information gleaned from the results of these scores would yield important and useful information to educators in regards to reading achievement and the effectiveness of instruction. However, some limitations are evident.

The high-stakes accountability movement calls for an assessment system that produces valid and reliable results that are standards-based and capable of prescribing

educational change that positively impact student learning (Good, Simmons, & Kame'enui, 2001). Although this group administered achievement test is considered to be both valid and reliable, the SAT9 fails to meet another important criteria because it is expensive and time consuming to administer and is only given once a year or in target grades such as 4th, 6th, and 9th grades. This creates a problem for using such tests as a diagnostic tool to aid in altering teaching strategies in order to positively impact student learning. Monitoring the progress of students once a year is not conducive to the prevention of reading failure. As mentioned earlier in this literature review, remediation is not the answer to solving the reading problem rather, it is early detection and prevention.

Dynamic Indicators Of Basic Early Literacy Skills

Aside from high stakes testing, educators are looking for a reliable, prevention-orientated, school-based assessment and intervention system in order to prevent early reading difficulty (Good, Simmons, & Kame'enui, 2001). One such assessment, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) published by the Institute for the Development of Educational Achievement from the University of Oregon, is used to identify early those children who may need additional instruction and support and to evaluate and modify instruction on an on-going basis to assure all children achieve (Good, Gruba & Kaminski, 2002). DIBELS measures were designed to assess students' early literacy skills as they change over time. The measures chart student growth, are easy and efficient to administer (each measure is a 1 minute fluency-based measure), can be administered frequently (each measure has several alternate forms), and are cost effective. DIBELS measures were not designed to be a comprehensive diagnostic reading

assessment. Rather, according to Good, Simmons, & Kame'enui, they are intended to "provide a fast and efficient indication of the academic well-being of students with respect to important early literacy skills" (2001, p. 8). Therefore, DIBELS measures can be considered much like curriculum-based measurement (CBM), an alternate form of assessment, which tracks student proficiency across core curriculum areas. DIBELS evaluate a set of early literacy skills identified in the literature as directly related to later reading competence—the Big Ideas, as mentioned earlier (Elliot, Lee, & Tollefson, 2001). The Big Ideas of Beginning Reading have directly influenced the DIBELS measures: (see chart below)

Big Ideas of Beginning Reading	**DIBELS Measure**
Phonological Awareness	Initial Sounds Fluency (ISF) Phoneme Segmentation Fluency (PSF)
Alphabetic Principal	Nonsense Word Fluency (NWF)
Fluency with Text	Oral Reading Fluency (ORF)

The following is a description of the DIBELS subtests or measures that are administered (as taken from the DIBELS 6th Edition Administration and Scoring Guide): (1) The Initial Sound Fluency (ISF) subtest is a standardized, individually administered measure of phonological awareness that assesses a child's ability to recognize and produce the initial sounds in an orally presented word. (2) The Letter Naming Fluency (LNF) subtest is a standardized, individually administered test that provides a measure of risk. Students are presented with a page of upper and lower case letters arranged in random order and are asked to name as many letters as they can. (3) The Phoneme Segmentation Fluency (PSF) subtest assesses a student' ability to segment three and four letter phoneme words into their individual phonemes fluently. (4) The Nonsense Word Fluency (NWF) subtest is a standardized, individually administered test of alphabetic

principle—including letter-sound correspondence and of the ability to blend letters into words in which letters represent their most common sounds. (5) The Oral Reading Fluency (ORF) subtest is a standardized, individually administered test of accuracy and fluency with connected text. (6) The Retell Fluency (RTF) subtest is intended to provide a comprehension check for the ORF assessment. (7) The Word Use Fluency (WUF) subtest is intended for students from fall of kindergarten through third grade. This subtest requires students to use the presented words in a coherent sentence. A benchmark goal is not established for the WUF because additional research is needed to establish this linkage to other big ideas of early literacy (Good & Kaminski, 2002).

According to the DIBELS manual, students are assessed three times a year using the prescribed subtests. Based on their performance, they are then placed in categories of Low Risk, Some Risk, and At-Risk as determined by the set DIBELS benchmark goals. The assessments are scored by imputing the raw scores into the specified database available through the DIBELS website. A charge of $1.00 per student is required for this data service.

According to Good, Gruba, & Kaminski, the minimum criteria for best practices in early literacy assessment must include the following: the assessment must be an effective prevention-oriented system that will reliably measure student growth on an ongoing basis; can predict success or failure on criterion measures (high stakes testing); and provide an instructional goal, that if met, will prevent reading failure (2001). DIBELS meets best practice criteria for measuring growth and development of early literacy skills in kindergarten and first grade. In addition, ongoing research funded by the Early Childhood Research Institute on Measuring Growth and Development, has

generated a large database demonstrating the psychometric adequacy of DIBELS. For example, by using DIBELS benchmarks in kindergarten and first grade, one can determine the level of skill that predicts risk by looking at benchmark goals. Even as early as kindergarten, one can determine, with a high degree of accuracy which children will have significant difficulty learning essential literacy skills unless additional instructional support is provided (Good, Gruba, & Kaminski, 2001).

The goal is to match students with the needed instructional support before a pattern of reading difficulty and failure is established. To do this, students are assessed using DIBELS. Individual students needing additional instruction are then identified by using the benchmark report which determines whether the student is Low Risk, Some Risk, or At-Risk and those who have met the benchmark goal, those who need strategic support, or intensive instructional support, respectively. According to student performance on the given DIBELS measures, a teacher can then direct specific instruction toward those particular students deficient in a specific skill area.

Standards-Based Achievement vs. Prevention-Oriented Diagnostic Assessments

All too often, assessment and intervention are treated as separate and unrelated activities. As mentioned earlier, the SAT9 and other popular achievement tests are time consuming and expensive to administer. In addition to these drawbacks, they are usually given once a year or only given in the target years, of the 4th, 6th and 9th grades, making it next to impossible to institute and monitor significant instructional changes or to facilitate effective interventions. The Council for Exceptional Children suggests for early intervention programs to be effective, they must be both intense and fast paced (Council for Exceptional Children, 1997). In addition, according to the US Dept. of Education

(1999), reading failure in the primary grades can be reduced to less than one in ten with quality early reading intervention programs. This is where the importance of a prevention-oriented diagnostic assessment such as DIBELS can prove to be effective.

Time and cost efficient, DIBELS can identify those children as early as kindergarten and fall of first grade those students who may have difficulty with reading. This early identification can initiate instructional change with frequent monitoring throughout the academic school year. Results of a study conducted by the Center for the improvement of Early Reading Achievement (1999) identified early reading intervention as a key factor in the successes of the most effective schools. Therefore, the best solution to the problem of reading failure is to allocate resources for early identification and prevention (Torgesen, 1998), such as implementing the use of a prevention-oriented assessment in the classrooms.

Purpose of Study

The purpose of the current study was to determine the concurrent validity of DIBELS, a type of curriculum based measurement, by correlating children's test performance to their obtained reading achievement scores on the Stanford Achievement Test-Ninth Edition, a standardized group administered achievement test—used by some states as a high-stakes test. The purpose was to determine the utility of DIBELS as a Curriculum Based Measurement, a prevention-oriented assessment, as it relates to high stakes testing success. The results of this study may encourage educators to use DIBELS as a classroom diagnostic tool to identify children's needs early and help prevent reading failure.

Hypothesis

It is hypothesized that a positive and significant correlation will exist between students' obtained DIBELS and SAT9 scores suggesting that DIBELS measures can be regarded as a valid inventory of students' early reading skills when compared to their reading achievement outcomes.

Method

Subjects

Seventy-nine students, 40 female and 39 male, from five first grade classrooms at a rural elementary school in southeastern Ohio, were involved in the current study. All subjects were Caucasian. The total enrollment for the elementary school is 574 students with 57% of those students receiving free or reduced lunch. The participants included were drawn from both regular and special education classrooms.

Among the archival data collected for the study, An entire classroom's (16 students) DIBELS PSF raw scores were not available to the investigator. In addition, a few students were absent during the administration of some subtests, which also resulted in missing data for a measure. The investigator, however, used the students' other available area raw score data measures for data analysis.

Instruments

In the spring, subjects were administered the SAT9 Primary 1, Form S using 1995 spring national normative data and in accordance to standardized procedures. The following DIBELS measures were also administered: Nonsense Word Fluency (NWF), Phoneme Segmentation Fluency (PSF), and Oral Reading Fluency (ORF) as suggested from the DIBELS administration manual.

Procedure

Permission was granted from the school principal in order to obtain the subjects' demographics, SAT9, and DIBELS scores from the school records. For the purpose of this study, the Pearson Correlation method was employed to explore the relationships between the SAT9 reading subtests and the DIBELS reading measures. The SAT9's Total Reading raw score and the Total Reading Cluster subtests Word Study, Word Reading, and Reading Comprehension raw scores were compared to the DIBELS' NWF, PSF, and ORF subtest raw scores. The raw scores were used to compare the assessments.

Results

The objective of this study was to investigate the concurrent validity between the SAT9 reading scores and the DIBELS reading measures. After gathering archival data from a small elementary school in rural southeastern Ohio, which includes the subjects demographics, SAT9 scores and DIBELS measures, the data was then entered into the Comprehensive Statistical Software Program (SPSS) version 10.0. The data were subjected to Descriptive Statistic analysis (see Figure 1). In addition, the Pearson Product Moment Correlation was used to explore the relationship between the students' SAT9 reading scores and their DIBELS measures (see Figure 2).

Results of the study indicated a significant and positive correlation between DIBELS' PSF measure and SAT9's Reading Comprehension ($r = .380$, $p = .002$), Word Study ($r = .5400$, $p = .0001$), and Total Reading ($r = .400$, $p = .001$) (see Figure 2). No significant correlation was indicated between DIBELS' PSF and SAT9's Word Reading

(r = .179, p = .161) (see Figure 2). Significant and positive correlations were also indicated between DIBELS' NWF and SAT9's Word Reading (r = .614, p = .000), Reading Comprehension (r = .611, p = .000), Word Study (r = .571, p = .000), and Total Reading (r = .639, p = .000) (see Figure 2). Finally, DIBELS' ORF is significantly and positively correlated with SAT9's Word Reading (r = .749, p = .000), Reading Comprehension (r = .728, p = .000), Word Study (r = .610, p = .000), and Total Reading (r = .740, p = 000) (see Figure 2). The study results indicated the concurrent validity between DIBELS measures and SAT9 reading scores range from poor (PSF and Word Reading) to clinically significant up to r = .749 and p = .0001 (see Figure 2).

Discussion

This study examined the relationship between DIBELS reading measures and SAT9 reading scores. The hypothesis of this study was that a positive and significant correlation will exist between students' obtained DIBELS and SAT9 scores suggesting that DIBELS measures can be regarded as a valid inventory of students' reading skills when compared to their reading achievement outcomes. The following question was examined in this study: What is the concurrent validity of the DIBELS reading measures in relation to students' SAT9 reading scores? The results of the study indicated that the concurrent validity between the DIBELS measures ranged from r = .380 and p = .002 to r = .749 and p = .000 with the exception of a non-significant correlation between DIBELS PSF and SAT9 Word Reading. This finding may be explained by the difference in the way a student learns to read. Meaning that, a student who has difficulty with hearing and sounding out phonemes in a given word may be a whole word reader or vice versa. The DIBELS ORF measure suggests being the best measure of concurrent validity in relation

to the SAT9 Reading scores. This finding is supportive with Crawford, Stieber and Tindal's research which cited a 1982 study in which investigators Deno, Mirkin, and Chiang established the criterion validity of reading aloud as a measure of general ability. The authors found that reading aloud was highly correlated with students' test performance ($r = .78$; $r = .80$) (2000). Crawford, Stieber and Tindal expanded on their study and found that 100% of second grade students in their study who read at least 72 correct words per minute passed the statewide reading test taken the following year. In the third-grade, 94% of the students reading less than 117 correct words per minute did not pass the statewide reading test taken during the same year (2000). Together, this information demonstrates the utility of a curriculum based measurement, one such as DIBELS, as a diagnostic and predictive tool for reading success in the classroom and on statewide achievement tests.

Variables not considered in this study might possibly lead to better validity outcomes. One such variable to consider is the homogeneity of the population. Due to the geographical location of the school, the study was unable to include racial and ethnic minorities and various socioeconomic statuses. Including such variables of diversity in race, ethnicity, and socioeconomic status and exploring these venues may provide additional insight to the study. Another variable to consider may be an exploration of the construct of the DIBELS measures and how they relate to the construct of the SAT9 subtests. In addition, the results of this study could have been underestimated due to the DIBELS PSF data not included from the classroom of 16 students, resulting in a smaller n for this measure in addition to the possibility that the omission of this data was not merely random.

Recommendations

Although the present study did not consider the variables discussed above, this study serves a purpose in that it demonstrates significant relationships between the DIBELS reading measures and SAT9 reading scores. Educators may use this study to evaluate the utility of DIBELS reading measures when used as a prevention-oriented, diagnostic reading tool in the classroom. Furthermore, this study may encourage further research on the construct of the two measures and their relationships, research investigations of DIBELS' predictive validity on high-stakes tests, and evaluating the use of DIBELS to measure instructional interventions in the classroom. In addition, the current study may be expanded by the exploration of additional variables such as race, ethnicity, and socioeconomic status in order to achieve a more heterogeneous sample.

References

Amrein, A. L., & Berliner, D. C., (2003). The effects of high-stakes testing on student motivation and learning. Educational Leadership, 60(5), 32-39.

Armbruster, B. B., Lehr, F., & Osborn, J., (2001). Put reading first: The research building blocks for teaching children to read. Jessup, MD: National Institute for Literacy and the Partnership for Reading.

Berk, R. A., (1997). Test review of the Stanford Achievement Test, Ninth Edition. From J.C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook [Electronic version]. Retrieved June 3, 2003 from the Buros Institute's Test Reviews Online website: http://www.unl.edu/buros.

Big ideas in beginning reading. Retrieved March 22, 2003 from http://reading.uoregon.edu/big_ideas/index.php.

Christie, K., (2001). Lagging Literacy. Phi Delta Kappan, 82(10), 729-732.

Council for Exceptional Children. (1997). Reading difficulties vs. Learning Disabilities. Council for Exceptional Children, 4(5), 1-5.

Crawford, L., Stieber, S., & Tindal, G. (2000). Using timed oral readings to predict students' performance on statewide achievement tests. University of Oregon, 1-22.

Elliot, J., Lee, S. W., & Tollefson, N., (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills—Modified. School Psychology Review, 30(1), 33-49.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). Dynamic indicators of basic early literacy skills (6th ed.): Administration and scoring guide. Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/.

Good, R. H., Gruba, J., & Kaminski, R. A. (1995). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.) Best Practices in School Psychology III, 679-700.

Good, R. H., Simmons, D. C., & Kame'enui, E. J., (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third grade outcomes. Scientific Studies of Reading, 5(3), 257-289.

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academi interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. School Psychology Review, 27(1), 45-57.

Haladyna, T. M., (1997). Test review of the Stanford Achievement Test, Ninth Edition, From J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook [Electronic version]. Retrieved June 3, 2003 from the Buros Institute's Test Reviews Online website: http://www.unl.edu/buros.

Harcourt Educational Measurement, Stanford 9 Overview (n.d.). Retrieved June 3, 2003, from http://www.hemweb.com/trophy/achvtest/sat9view.htm.

High-Stakes tests don't improve achievement. (2003). Education USA, 45(1), (1,7).

International Literacy Network. Facts and Questions. Retrieved May 3, 2003 from http://www.theiln.org/faq.htm.

Kaminski, R. A., & Good, R. H., (1996). Toward a technology for assessing basic early literacy skills. School Psychology Review, 25(2), 215-228.

Lyon, R., (1997). Report on learning disabilities research. National Institute of Child Health and Human Development & Nation Institutes of Health. Proceedings of a testimony given by Dr. Reid Lyon before the Committee on Education and the Workforce in the U.S. House of Representatives. Retrieved from http://www.ldonline.com on June 3, 2003.

No child left behind act of 2001: Reauthorization of the Elementary and secondary education act legislation and policies. Retrieved from http://www.ed.gov/offices/OESE/esea/ on June 3, 2003.

Stollar, S. A. (2002). Preventing reading failure through early literacy assessment and instruction. The Ohio School Pscyhologist, 47(2), 35-37.

Taylor, B., Pearson, P.D., Clark, K., & Wapole, S. (1999). Beating the odds in teaching children to read. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. American Educator, Spring/Summer, 1-18.

U.S. Department of Education. (1999). Start Early, Finish Strong. Washington, D.C.: U.S. Department of Education.

Appendix A

Figure 1

Descriptive Statistics

	Mean	Std. Deviation	N
sex	1.5190	.50283	79
years in months	87.8101	6.08081	79
WORREAD	22.4937	6.65423	79
READCOM	31.1392	7.78186	79
WORDSTUD	28.5063	6.28566	79
total reading score	82.0127	20.01569	79
PSF	46.2381	10.24178	63
NWF	61.4675	30.20743	77
ORF	57.2692	38.63942	78

Figure 2

Correlations

		sex	years in months	WORREAD	READCOM	WORDSTUD	total reading score	PSF	NWF	ORF
sex	Pearson Correlation	1	.238*	-.189	-.297**	-.380**	-.304**	-.136	-.123	-.109
	Sig. (2-tailed)	.	.035	.096	.008	.001	.006	.289	.288	.342
	N	79	79	79	79	79	79	63	77	78
years in months	Pearson Correlation	.238*	1	-.132	-.197	-.257*	-.190	-.263*	-.176	-.195
	Sig. (2-tailed)	.035	.	.247	.081	.022	.094	.038	.125	.087
	N	79	79	79	79	79	79	63	77	78
WORREAD	Pearson Correlation	-.189	-.132	1	.810**	.728**	.903**	.179	.614**	.749**
	Sig. (2-tailed)	.096	.247	.	.000	.000	.000	.161	.000	.000
	N	79	79	79	79	79	79	63	77	78
READCOM	Pearson Correlation	-.297**	-.197	.810**	1	.840**	.961**	.380**	.611**	.728**
	Sig. (2-tailed)	.008	.081	.000	.	.000	.000	.002	.000	.000
	N	79	79	79	79	79	79	63	77	78
WORDSTUD	Pearson Correlation	-.380**	-.257*	.728**	.840**	1	.911**	.540**	.571**	.610**
	Sig. (2-tailed)	.001	.022	.000	.000	.	.000	.000	.000	.000
	N	79	79	79	79	79	79	63	77	78
total reading score	Pearson Correlation	-.304**	-.190	.903**	.961**	.911**	1	.400**	.639**	.740**
	Sig. (2-tailed)	.006	.094	.000	.000	.000	.	.001	.000	.000
	N	79	79	79	79	79	79	63	77	78
PSF	Pearson Correlation	-.136	-.263*	.179	.380**	.540**	.400**	1	.288*	.239
	Sig. (2-tailed)	.289	.038	.161	.002	.000	.001	.	.022	.060
	N	63	63	63	63	63	63	63	63	63
NWF	Pearson Correlation	-.123	-.176	.614**	.611**	.571**	.639**	.288*	1	.828**
	Sig. (2-tailed)	.288	.125	.000	.000	.000	.000	.022	.	.000
	N	77	77	77	77	77	77	63	77	77
ORF	Pearson Correlation	-.109	-.195	.749**	.728**	.610**	.740**	.239	.828**	1
	Sig. (2-tailed)	.342	.087	.000	.000	.000	.000	.060	.000	.
	N	78	78	78	78	78	78	63	77	78

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).