1-1-2009

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A Comparison of Scoring Systems

Ashley E. Barr
ashleybarrwv@gmail.com

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A Comparison of Scoring Systems

Thesis submitted to the
Graduate College of
Marshall University

In partial fulfillment of
the requirements for the degree of
Educational Specialist
In School Psychology

by

Ashley E. Barr

Fred Jay Krieg, Ph.D., Committee Chairperson
Sandra S. Stroebel, Ph.D.
Edna M. Meisel, Ed.D.

Marshall University

May 2009

ABSTRACT

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A Comparison of Scoring Systems

By Ashley E. Barr

The purpose of this study was to compare scoring systems of the Dynamic Indicators of Basic Early Literacy Skills, commonly known as DIBELS. Currently, there are two systems for scoring the DIBELS assessments. The first system is a paper and pencil approach and the second system is electronic and utilizes a handheld palm-pilot. This study determined whether the electronically scored DIBELS produces the same scores as the hand-scored DIBELS. It was hypothesized that the electronically scored DIBELS will yield significantly different scores than the hand-scored DIBELS. Median scores obtained from the electronic palm-pilot were compared to median scores obtained from the paper and pencil scoring method. A total of 82 first grade students at North Elementary School in Morgantown, WV were included in this study. Results were analyzed using the t-test for independent samples statistical method. No statistically significant differences were found between the DIBELS scoring systems.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First, I would like to thank my entire family for their patience, love, and support over the past three years and for providing me with the inspiration and encouragement I needed to complete this challenging task and accomplish my goal of becoming a school psychologist. I would like to express gratitude to my field supervisor, Lisa Gainer, who assisted me in finding a school that would be open and willing to let me collect data during school hours and for helping me organize my thoughts throughout the writing process. I would also like to extend my gratitude to Dr. Meisel for the assistance with the data analysis and guidance with the overall process of thesis development. To Dr. Stroebel, I would like to express my gratitude for helping me obtain the appropriate approval for the study, for keeping me aware of deadlines, and for keeping me motivated to complete this process. Finally, I would like to extend my deep appreciation to Dr. Krieg for not only his guidance throughout this program as a professor and as a mentor, but also for his insight and dedication in helping me complete this study.

List of Tables

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A Comparison of Scoring Systems

CHAPTER I

LITERATURE REVIEW

Recent research has revealed the importance of assessing the beginning reading skills of young children in hopes of promoting reading success in their futures. The assessment and intervention of reading literacy skills in young children is a crucial step for the prevention of early reading failure and essential for the creation of a strong reading foundation.

The National Assessment of Education Progress (NAEP) has documented high rates of reading failure among fourth and eighth graders (Nation's Report Card, 2007). These rates have been rising over the past years and have caused concern for the nation's educators and political decision makers. In areas of poverty as well as in middle class societies, children are failing to read at even the most basic levels, especially, minority children (Nation's Report Card, 2007). Failure means that most of these children cannot function at grade level reading standards and cannot independently keep up in their school work. Thus, the need to detect and prevent reading problems as early as kindergarten is of vital importance for our society.

The No Child Left Behind Act (NCLB, 2001) and the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) have influenced educational goals by requiring that the academic experiences of all students be enhanced by the use of high quality research-based instructional methods (Cusumano, 2007). The ultimate goal is for all students regardless of gender, race, socioeconomic status, or identified disabilities to be able to reach proficiency in all academic areas, with the most emphasis in the area of reading (Cusumano, 2007). The importance of the early identification of students who fall off track in meeting these goals is evident. Early identification required an assessment instrument that monitored students'

acquisition of skills in areas of basic academic achievement such as reading. Instruments for collecting this data must be sensitive to small changes in skill acquisition, must not confound future data due to frequent administration, and must be quick and easy to administer (Cusumano, 2007). Test instruments that address all of these needs are known as Curriculum Based Measures (CBM).

One Curriculum Based Measurement that has quickly gained popularity in the field of education is the Dynamic Indicators of Basic Early Literacy Skills, or DIBELS. Unlike other early reading assessments, DIBELS are quick, cheap, and relatively easy to use and have been aggressively promoted by the federal government and consultants to the Reading First program. Because the federally funded Reading First program was the first to use the DIBELS system successfully in schools and endorsed its use within the three-tiered process, the DIBELS system seems to have a competitive edge over other reading assessments. DIBELS is a reliable and valid measurement tool in predicting which children may have reading difficulties later (Good & Kaminski, 2003). These brief, simple assessments are measures of the critical skills that underlie early reading success and have been extensively researched and used to predict how well children are likely to be doing in reading comprehension by the end of the third grade and beyond. The DIBELS assessments are designed to measure early reading skills that are critical in determining reading success such as initial sound fluency, letter naming fluency, phoneme segmentation fluency, nonsense word fluency, oral reading fluency, retell fluency, and finally word use fluency.

In West Virginia, DIBELS assessments are given to all students in a class three times a year and scores reveal those who are at-risk for reading failure. Teachers are then able to intervene with those students by working on specific reading literacy skills and giving progress-

monitoring assessments to help these children overcome their reading deficits. DIBELS scores are used to help predict the scores on reading achievement tests given at the end of the school year as well as for grouping students in the classroom based on specific learning needs (Good & Kaminski, 2003). These scores are very sensitive to instruction and can be measured over short time periods (Good & Kaminski, 2003). Because these assessments have been used extensively in predicting students' reading success and failure and for grouping students for learning, it is critical for the scoring systems of DIBELS to be valid and reliable.

Effectiveness of Curriculum Based Measurement

Teachers and school psychologists are utilizing an increasingly popular form of alternative assessment, known as Curriculum-Based Measurement (CBM), to monitor student progress in the classroom (Stecker, Fuchs & Fuchs, 2005). In the mid-to-late 1970s during the time of the original passage and implementation of IDEA, Stan Deno and colleagues at the Institute for Research on Learning Disabilities (IRLD) at the University of Minnesota wanted to develop a simple and efficient, but technically sound instrument that could be used to help special educators track the growth of their students' basic skills. When first developing the CBM, their purpose was to assist special educators in using progress-monitoring data to make meaningful decisions about student progress and to improve the quality of instructional practices (Stecker, Fuchs & Fuchs, 2005). The past two decades of research reveals that CBM are used in a variety of ways from screening and identifying students in need of special education to monitoring progress and planning instruction in the general education classroom (Stecker, Fuchs & Fuchs, 2005). Keeping the original intent of CBM in mind, the hope was for teachers to be able to respond quickly to students' poor patterns of performance and change instruction in order to enhance student achievement.

Stecker, Fuchs and Fuchs (2005) reviewed studies and examined the efficacy of CBM as an assessment methodology for enhancing student achievement. They looked at studies where teachers used CBM to monitor student progress and make instructional decisions in the areas of math and reading. They were able to draw several conclusions from their investigations. First, teachers had an effect on student achievement when they used progress monitoring data to modify their instruction; however, frequent progresses monitoring alone without the use of instructional modification appeared to have no effect on improving student achievement. Second, teachers appeared to be more responsive to student needs when they used data-based decision rules to interpret graphed CBM data. Third, computer applications helped teachers utilize the decision rules and incorporated a goal-raising feature that also stimulated student growth. Overall, the use of computer applications led to teacher satisfaction of the CBM procedures. In the general education environment, teachers who used their own methods for progress monitoring and instruction did not have as great of an effect on academic growth across students of varying academic histories compared to the teachers who used class wide CBM data and recommendations for instructional planning, which incorporated the implementation of PALS, or Peer-Assisted Learning Strategies. (Stecker, Fuchs & Fuchs, 2005).

In 2004, Madelaine and Wheldall reviewed a large body of research published within the last ten years on CBM of reading, in particular, on the technical characteristics of the oral reading fluency measure. Included in this review was Marston's (1989) research, where ORF emerged as the most valid indicator of overall reading performance. Other studies repeatedly showed that ORF was also a valid measure of reading comprehension (Deno, Mirkin & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993). The use of a maze task emerged as the main alternative to ORF (Fuchs & Fuchs, 1992). The maze task requires students to read a

passage in which every nth word is deleted. When students come to a deleted word, they must choose an answer from a multiple choice item, which contains the correct word and two distracters. It was reported that the main advantage of using the maze task over ORF is that the maze task does not require the teacher or other professional to administer the test to students individually. The maze task can be administered by a computer or to groups of students and is therefore less time consuming than a typical test of ORF (Fuchs, 1992, 1998). Although the maze task has these advantages, the research consistently reveals that ORF is a more reliable and valid indicator of reading performance than a maze (Faykus & McCurdy, 1998).

An overview of CBM by Cusumano (2007) reviewed the research on the DIBELS measures as an assessment of early literacy. Strong psychometric properties were documented for all of the DIBELS subtests, including the Nonsense Word Fluency and Oral Reading Fluency subtests which were utilized in the present study. The AIMSweb Maze task was another example of CBM discussed in the review. The AIMSweb Maze assessment monitors reading comprehension and researchers have provided documentation of its validity, reliability, and sensitivity for monitoring student growth (Shin, Deno, & Espin, 2000).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

The development of DIBELS was based upon the measurement procedures used for Curriculum-Based Measurement (CBM) by Deno and colleagues through the Institute for Research on Learning Disabilities at the University of Minnesota in the 1970s-80s (e.g., Deno, 1985; Deno & Fuchs, 1987; Deno & Mirkin, 1977; Shinn, 1989). DIBELS are similar to CBM in that they were developed to be economical and efficient indicators of a student's progress toward achieving a general outcome. Initially, DIBELS materials were developed to be linked to the local curriculum like other CBM (Kaminski & Good, 1996). Current DIBELS measures are

generic and do not base their content on any specific school's curriculum. General Outcome Measurement (GOM) is the term being used to describe the generic CBM methodology (Fuchs & Deno, 1994).

General Outcome Measures (GOMs) like DIBELS are different from other commonly used formative assessment approaches. When assessing a child's progress in the curriculum, the most common formative assessment approach that teachers use is called mastery measurement. One example of mastery measurement is a test given at the end of a unit. Teachers teach certain skills and then test to see if the students master these skills. They continue to teach the next skills in the sequence and test for mastery of those skills. As they move from test to test, both the type and difficulty of the skills assessed change; therefore scores from different times in the school year cannot be compared. Mastery-based formative assessment such as end of unit tests addresses whether the student has learned the content taught while GOMs like DIBELS are designed to see if students are learning and making progress towards a goal (Kaminski, Cummings, Powell-Smith & Good, 2008).

Reliability

Good, Kaminski, Simmons, and Kame'enui (2001) reviewed the use of DIBELS and Curriculum-Based Measurement Reading within an outcomes-driven model. This model was prevention-oriented and designed to detect those students at-risk for reading problems early in an effort to help them achieve successful reading outcomes by the end of the third grade. The model included a conceptual foundation of crucial early literacy skills for assessment and instruction which was based on empirically sound outcomes for each early literacy skill. This model was created using reliable and valid measures of the DIBELS assessment tool to document growth toward reading outcomes. It took data generated from these measures and provided a set of steps

for developing a curriculum at the individual and systems level. Some of the data used to develop the curriculum included the following: Onset Recognition Fluency measure had an alternate-form reliability of 0.72, the Phoneme Segmentation Measure had a 2-week alternate-form reliability of .88, the one-month alternate-form reliability of Nonsense Word Fluency was .83, and Letter Naming Fluency had a one-month alternate-form reliability of .88. The big ideas in beginning reading including phonemic awareness, alphabetic principle, accuracy and fluency with connected text, and high stakes reading outcomes were thoroughly described and steps for following the model were given using illustrations to guide the reader. This study utilized the paper and pencil scoring method to obtain data for the DIBELS measures.

Validity

A study examining the concurrent validity and diagnostic accuracy of the DIBELS assessment with the Comprehensive Test of Phonological Processing (CTOPP) was carried out by Hintze, Ryan, and Stoner (2003). The CTOPP is an assessment of phonological awareness, phonological memory and rapid memory. This study included 86 kindergarten students from a midsized city in Northwestern Massachusetts. Six graduate students were trained to collect data from both the DIBELS and CTOPP tests. Data was collected using the paper and pencil scoring method. Correlation coefficients were obtained for the DIBELS measures and subtest scores and also for the CTOPP scores. After examining the coefficients, DIBELS measures were shown to be strongly correlated with most of the subtests and composite scores of the CTOPP. The subtests of the CTOPP that measured phonological awareness and memory included Elision, Blending Words, Sound Matching and Nonsense Word Repetition and these were most strongly correlated with DIBELS subtests of Initial Sound Fluency (ISF), r = .52, .51, .51, and .44, respectively and Phonemic Segmentation Fluency (PSF), r = .47, .63., .25, and .33 respectively.

The Letter Naming Fluency (LNF) subtest of DIBELS correlated strongly with subtest and composite scores of the CTOPP that represented phonemic awareness (r = .45, .38, .53, and .44) and memory, as well as rapid naming abilities (r = .59, .59, .43, .53, .52, and .58).

In order to look at the diagnostic accuracy of DIBELS, studies were carried out using DIBELS as the predictor variables and the CTOPP as the criterion measure. The author's use of cut-scores resulted in extremely high levels of sensitivity and low levels of specificity for both the ISF and PSF tasks of DIBELS. This meant that the students tested with these two DIBELS measures were correctly classified according to their CTOPP composite scores. For example, the ISF task of the DIBELS is quite sensitive to both the Phonological Memory Composite (PMCom) and the Phonological Awareness Composite (PACom) scores of the CTOPP (1.00 and 0.91, respectively). However, the specificity of the ISF task appears to be low (.39 and .36 for PACom and PMCom, respectively). In conclusion, the moderate to strong correlations found between these two instruments provide evidence that they are measuring a similar construct and can both be used for assessing a child's skills in phonological awareness.

Statement of the Problem

Currently, there are two widely used systems for scoring the DIBELS assessments. The first system utilizes a paper and pencil approach, using published scoring materials and recording the scores by hand. The second system is electronic and utilizes a handheld palm-pilot, which makes the correct calculations and stores the data in a computer system for future reference. Are the newly developed computerized scoring programs resulting in the same scores as the paper and pencil scoring system?

The purpose of this study was to determine whether or not the electronically scored DIBELS produces the same scores as the hand-scored DIBELS. Such results are important

because currently there are no studies concerning whether these recently developed computerized scoring programs are producing the same results as the paper and pencil scoring system by which the DIBELS assessments were normed. The Wireless Generation Company, namely Mobile Classroom Assessment or MCLASS, who is producing and selling the computerized scoring systems to schools, are claiming that the scores provided by their computerized instruments are the same as the scores obtained from the pencil and paper scoring procedure. However, there is no research to support their claim.

Statement of Hypothesis

It is hypothesized that when a sample of first grade students are given the Oral Reading Fluency (ORF) and Nonsense Word Fluency (NWF) benchmark assessments of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scored by a computerized palm-pilot and then by the hand scoring system, the electronically scored DIBELS will yield significantly different scores than the hand-scored DIBELS. The null hypothesis is that there will be no significant difference in scores between the hand-scored DIBELS and the electronically scored DIBELS.

Definition of Terms

The following terms important to this study are defined:

1) The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are reading assessment tools used to measure early reading skills in children. Research has revealed that these assessments predict reading comprehension abilities of children up to the end of the third grade and beyond.

2) DIBELS Oral Reading Fluency (ORF) is a standardized, individually administered test of accuracy and fluency with connected text.

3) DIBELS Nonsense Word Fluency (NWF) is a standardized, individually administered test of the alphabetic principle, including letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds.

3) Electronically scored DIBELS means that the assessment is scored by a small computerized handheld device, or palm-pilot, where the teacher enters data from the assessment and the computer system produces the score.

4) Hand-scored DIBELS means that the assessment will be scored using a pencil and paper approach. All of the calculations will be done by hand to produce the score.

CHAPTER II

METHOD

Research Design

The research design for the current study was quasi-experimental due to sampling procedures. The sample for this study was considered a sample of convenience as the participants were not chosen at random and came from only one school district. The scoring method was the independent variable with factors being the pencil and paper method and the palm-pilot method. The dependent variable was the resulting scores obtained from each method. Data collected from the study showed whether or not there was a difference among the scoring methods.

Participants

All students (ages 6-7 years) in the first grade, were administered the DIBELS winter benchmark assessments in Oral Reading Fluency (ORF) and Nonsense Word Fluency (NWF). They were selected from a school (North Elementary) within the Monongalia County School System in Morgantown, West Virginia. Students came from low to high socioeconomic backgrounds. Consent of the school system and the building principal was obtained in order to use the winter benchmark DIBELS ORF and NWF data and to administer and score additional ORF and NWF assessments by the paper and pencil scoring method. No parental permission was needed as this study was only interested in the numbers produced by the different scoring methods. Names of students were not included in the study. Four first grade classrooms were included in the study. Classrooms contained approximately 20 to 25 students each. Eighty-two first grade students were included in this study. Due to only one school being used in this study, the sample was not fully representative of the general population. North Elementary School's

population consists of a large number of international students. Out of 674 students, 156 speak 39 different languages and are from over 40 countries. Approximately 25% of the population consists of children who speak English as a second language. Out of 674 students, 72.5% are White/Non-Hispanic, 10.5% are Black/Non-Hispanic, 3.3% are Hispanic, 0.9% are Native Americans/Alaskan Natives, and 12.8% are Asian/Pacific Islanders. At North, the demographics are slightly skewed as there are more students at the school with highly educated parents compared to the average student populations in other county schools. The percentage of students receiving free or reduced lunch at North is 28%.

Instruments

Data was collected using a published instrument called the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS are a set of standardized, individually administered measures of early literacy development. They are designed to be short (one minute) fluency measures used to regularly monitor the development of pre-reading and early reading skills. Results can be used to evaluate individual student development as well as provide grade level feedback toward validated instructional objectives.

A series of studies have investigated the reliability, predictive validity, concurrent validity, construct validity, and item sensitivity of DIBELS. According to a study by Good, Gruba, and Kaminski (2001) alternate-form reliability for DIBELS ranged from .65 to .93 while the reliability of onset recognition fluency was a low .65. Concurrent criterion-related validity of DIBELS measures with other standardized measures of early literacy skills ranged from .36 for the onset recognition fluency to .81 for the letter naming fluency. Looking at outcomes 1 year later, the predictive validity correlations ranged from .36 to .82 and it was reported that all measures showed sensitivity to growth over time. The benchmark assessments of DIBELS

include *Initial Sound Fluency or Onset Recognition Fluency* (Kindergarten), *Letter Naming Fluency* (Kindergarten to Grade 1), *Phoneme Segmentation Fluency* (Mid-Kindergarten to End Grade 1), *Nonsense Word Fluency* (Mid-Kindergarten through Beginning Grade 2), and *Oral Reading Fluency* (Mid-Grade 1 to Grade 6). Oral Reading Fluency and Nonsense Word Fluency were the only two assessments administered in this study. These measures provided numerical test scores.

The DIBELS Oral Reading Fluency (ORF) is a standardized, individually administered test of accuracy and fluency with connected text. It is intended for most children from mid first grade through third grade. Student performance is measured by having students read a passage aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The number of correct words per minute from the passage is the oral reading fluency rate.

The ORF measure was chosen because it is the most researched measure of the seven DIBELS assessments. Also, ORF is currently the only measure being used in the upper grades (3-6). A number of studies on oral reading fluency have confirmed the technical adequacy of CBM Reading procedures in general (Good & Jefferson, 1998; Tindal, Marston, & Deno, 1983). Measures of oral reading fluency have been found to be indicators of future reading achievement (Shinn, 1997). A large body of research has validated CBM ORF measures as an excellent overall measure of reading achievement (Marston, 1989; Shinn, 1989; 1998). Further information regarding the reliability and validity of ORF can be found at http://dibels.uoregon.edu.

The DIBELS Nonsense Word Fluency (NWF) measure was also chosen for this study, although there is not as much research on NWF compared to the other DIBELS measures. The

NWF measure is a standardized, individually administered test of the alphabetic principle - including letter-sound correspondence in which letters represent their most common sounds and of the ability to blend letters into words in which letters represent their most common sounds (Kaminski & Good, 1996). It is intended for most children from mid Kindergarten through the beginning of second grade. Students read a sheet of paper with randomly ordered vowel-consonant and consonant-vowel-consonant nonsense words (e.g., sig, rav, ov) and are asked to verbally produce the individual letter sound of each letter or verbally produce, or read, the whole nonsense word. In January of first grade, the NWF has a one-month alternate-form reliability of .83 (Good, Kaminski, Shinn, Bratten, Shinn, Laimon, Smith, & Flindt, 2004). The DIBELS NWF also has concurrent criterion-validity with the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness Cluster with a score of .36 in January and .59 in February of first grade (Good et al., 2004).

Procedure

Permission and approval to conduct the study was granted by the Marshall University Graduate College Institutional Review Board (IRB) prior to beginning the study. Once permission was granted by the IRB, a letter requesting permission to collect the DIBELS data at North Elementary School during school hours was signed by the building principal. Data was collected during the first and second weeks of February, 2009. A reading specialist at North Elementary School and the first grade teachers collected benchmark data using the computerized palm-pilots with first grade students only. The day after collecting benchmark data using the palm-pilots, the researcher began collecting additional data using the paper and pencil scoring method. The paper and pencil data collection was completed prior to any student receiving Tier II or Tier III reading intervention. The researcher downloaded the DIBELS ORF and NWF

Progress Monitoring First Grade Administration and Scoring Booklets provided by the official DIBELS website in order to administer and score the measures using the paper and pencil method. The researcher used the ORF Progress Monitoring Stories six, nine, and twelve in order to obtain a median ORF score and the NWF Progress Monitoring form ten in order to obtain a numerical test score to compare to the computerized scoring method. Statistical analyses were conducted following the data collection.

CHAPTER III

RESULTS

Data Analysis

The purpose of this study was to determine whether or not the electronically scored DIBELS produces the same scores as the hand-scored DIBELS. Interval data, represented by numerical test scores from the DIBELS assessments was collected and the t-test for independent samples statistical method was used to analyze this data. The resulting means of the two scoring methods were compared. The 2007 Microsoft Excel Program software was utilized in the data analysis procedures.

The research hypotheses included: 1) When a sample of first grade students are given the Oral Reading Fluency (ORF) benchmark assessment of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scored by a computerized palm-pilot and then by the hand scoring system, the electronically scored DIBELS will yield significantly different scores than the hand-scored DIBELS; and 2) When a sample of first grade students are given the Nonsense Word Fluency (NWF) benchmark assessment of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scored by a computerized palm-pilot and then by the hand scoring system, the electronically scored DIBELS will yield significantly different scores than the hand-scored DIBELS.

As shown in Table 1 below, there was not a statistically significant difference between the ORF assessment scored by the palm-pilot and the ORF assessment scored by the paper and pencil method. Also, there was not a statistically significant difference between the NWF assessment scored by the palm-pilot and the NWF assessment scored by the paper and pencil method.

Table 1

Comparison of Scoring Systems

Type of assessment	**Mean of palm-pilot**	**Mean of paper/pencil**	**t-test**	**Probability level attained**
Oral Reading Fluency (ORF)	51.5	54.2	0.446	0.656
Nonsense Word Fluency (NWF)	65.4	64.7	0.125	0.901

*Significance attained at $p<0.05$

CHAPTER IV

DISCUSSION

This study examined whether or not the DIBELS assessments scored by an electronic system or palm-pilot were producing the same scores as the DIBELS assessments scored by hand, or the paper and pencil scoring method. It was hypothesized that the electronically scored DIBELS would produce significantly different scores than the hand-scored DIBELS. The resulting means from each scoring method were analyzed using the t-test for independent samples statistical method.

Research by Stecker, Fuchs and Fuchs (2005) points to the fact that Curriculum Based Measurement (CBM) is an important tool for monitoring student progress and growth within the classroom and enhancing student achievement. Also, studies by Cusumano (2007) show that a CBM like the DIBELS assessments, in particular Oral Reading Fluency, are reliable and valid in predicting students' overall reading achievement. Since the DIBELS assessments are being heavily promoted across the state of West Virginia for predicting students' reading success and failure and for grouping students for learning, it is critical for the scoring systems to be reliable and valid. Although the DIBELS assessments were normed using the paper and pencil method, many schools have recently purchased software systems, utilizing a palm-pilot, that claim to provide a faster and easier approach to scoring. Thus, there was a need for the current study which attempted to examine whether these newly developed electronic scoring systems are resulting in the same scores that are produced by the hand scoring method.

The results of this study showed that there were no statistically significant differences between the DIBELS assessments that were scored by the paper and pencil scoring method and the DIBELS assessments that were scored by the palm-pilot scoring method. These results are

important because teachers who have discontinued the use of the paper and pencil scoring method in exchange for the electronic scoring method, which is believed to be more time efficient and easier to use, are obtaining the same scores when assessing their students. A further investigation into whether or not the electronic scoring method is indeed a faster, easier approach to scoring would be an additional reason for utilizing the palm-pilot scoring method instead of the paper and pencil scoring method.

Limitations and Delimitations

Limitations of this study include the fact that the researcher was involved in collecting the data using the paper and pencil scoring method. This resulted in experimenter bias as the researcher could have influenced the outcome of the study to support the research hypotheses. Also, different people were involved in the data collection procedures and could have scored the DIBELS assessments differently. For example, the reading specialist and first grade teachers at North collected the data using the palm-pilot scoring method while the researcher collected the data using the paper and pencil scoring method. There was also a flaw in the research design. All students were scored by the electronic scoring system before being scored by the paper and pencil scoring method. The researcher should have counterbalanced the administration of the electronic and paper and pencil scoring methods to reduce any type of practice effect. Delimitations used to narrow this study for researchability include the use of only one school, and the use of only four first grade classrooms in the sample of participants, which may not be fully representative of the general population.

References

Cusumano, D. L. (2007). Is it working?: An overview of curriculum based measurement and its use for assessing instructional, intervention, or program effectiveness. *The Behavior Analyst Today, 8*(1), 24-34.

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

Deno, S.L. & Fuchs, L.S. (1987). Developing curriculum-based measurement systems for databased special education problem solving. *Focus on Exceptional Children, 19*(8), 1-15.

Deno, S.L. & Mirkin, P.K. (1977). *Data-based program modification: A manual*. Reston VA: Council for Exceptional Children.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.

DIBELS Data System Web site: https://dibels.uoregon.edu/.

Faykus, S. P., & McCurdy, B. L. (1998). Evaluating the sensitivity of the maze as an index of reading proficiency for students who are severely deficient in reading. *Education and Treatment of Children, 21*, 1-21.

Fuchs, L. S. (1998). Computer applications to address implementation difficulties associated with curriculum-based measurement. [Electronic version]. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 89-112). New York: The Guilford Press.

Fuchs, L.S. & Deno, S.L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 6*(1), 15-24.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21,* 45-58.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20-28.

Good, R. H., III, Gruba, J., & Kaminski, R. A. (2001). Best practices in using dynamic indicators of basic early literacy skills (dibels) in an outcomes-driven model. *Best Practices in School Psychology, IV,* (pp. 679-700). Washington, DC: National Association of School Psychologists.

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on Curriculum-Based Measurement validity. [Electronic version]. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 61-88). New York: Guilford.

Good, R. H. III, & Kaminski, R. A. (Eds.). (2003). Dynamic indicators of basic early literacy skills (6th ed.). Colorado: Sopris West.

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., et al. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development.* (Tech. Rep. No. 7). Eugene: University of Oregon.

Good, R. H., III, Kaminski, R. A., Simmons, D., & Kame'enui E. J. (2001). Using dynamic indicators of basic early literacy skills (dibels) in an outcomes-driven model: Steps to reading outcomes. *Oregon School Study Council, 44,* 2-26.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review, 32*(4), 541-556.

Individuals with Disabilities Education Improvement Act of 2004. 20SS611–614 (2004).

Jenkins, J. R., & Jewll, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59,* 421-432.

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

Kaminski, R., Cummings, K.D., Powell-Smith, K.A., & Good, R.H.III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills for formative assessment and evaluation. In A. Thomas and J. Grimes (eds.) *Best practices in school psychology V*. Bethesda, MD: National Association of School Psychologists.

Madelaine, A., & Wheldall, K. (2004). Curriculum base measurement of reading: Recent advances. *International Journal of Disability, Development and Education, 51*, 57-82.

Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18- 78), New York: Guilford Press.

Nation's Report Card. (2007). Available from the National Assessment of Education Progress Web site, http://nces.ed.gov

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

Shin, J., Deno, S.L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34(3),* 164-173.

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

Shinn, M. R. (1997). *Instructional Decision Making Using Curriculum-Based Measurement*. Unpublished workshop materials.

Shinn, M. R. (Ed.) (1998). *Advanced Applications of Curriculum-Based Measurement*. New York: The Guilford Press.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795-819.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.