EXPLORING CONSISTENCY IN DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS NEXT ORAL READING FLUENCY PASSAGES FOR EDUCATIONAL DECISION MAKING

Thesis

Submitted to

The School of Education and Allied Professions of the UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for The Degree of Educational Specialist in School Psychology

By Debra Joyce Gillespie

Dayton, Ohio

August, 2012
EXPLORING CONSISTENCY IN DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS NEXT ORAL READING FLUENCY PASSAGES FOR EDUCATIONAL DECISION MAKING

Name: Gillespie, Debra J.

APPROVED BY:

__________________________________
Sawyer A. Hunley, Ph.D.
Advisory Committee Chairman
Associate Professor
Department of Counselor Education & Human Services

_________________________________
Susan Davies, Ed.D.
Committee Member
Assistant Professor
Department of Counselor Education and Human Services

____________________________________
Elana R. Bernstein, Ph.D.
Committee Member
Clinical Faculty
Department of Counselor Education and Human Services
ABSTRACT

EXPLORING CONSISTENCY IN DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS NEXT ORAL READING FLUENCY PASSAGES FOR EDUCATIONAL DECISION MAKING

Name: Gillespie, Debra Joyce
University of Dayton

Advisor: Dr. Sawyer Hunley

The purpose of this thesis was to explore a curriculum based measurement procedure and its reliability for practical applications in the educational setting. Twenty third grade oral reading passages were selected from the Dynamic Indicator of Basic Early Literacy Skills (DIBELS) system and used to assess reading progress of three third grade classes from different school districts in southwestern Ohio.

The DIBELS system includes a comprehensive assessment system that is validated through research and assesses early reading skill development from kindergarten through sixth grade. The standard procedure of using the one-minute median probe of a three probe administration was compared to the alternative procedure of using only the first of three probes. The correlational analysis of single first
probescores was positive and highly significant with median score probes $r (537) = .96$, $p < .01$. This study’s discussion included the innovation of progress monitoring tools, the utility of such, legal enactments, a decision making framework in reading for progress monitoring use with the K-12 school system and exploration of a more efficient and reliable solution.
In loving memory of my parents, Francis and Bernice; and my late husband, Tommie, whose love provided the basis for growth.

To my children, Jeremy, Matthew, and Holly: I have witnessed the power of Christ and of education in shaping your developing minds into caring, conscientious, and accomplished young adults.
ACKNOWLEDGEMENTS

I thank the distinguished Dr. Sawyer Hunley for allowing me to participate in such an important research project and for her invaluable advice in planning, writing, and implementing the research. I have leaned on her as an academic advisor, a committee chairperson, and a friend. I also wish to thank the other distinguished members of my committee: Dr. Susan Davies, whose energy abounds as a teacher and researcher; and Dr. Elana Bernstein who graciously agreed to assist in advising and in offering technical formatting for this research project. Special thanks to Dr. Laura Berry Kuchle for her diligence in reading and rereading the literature review. She provided invaluable insight into the inconsistencies and illogical conclusions inherent in first drafts.

I wish to thank Mrs. Joel Mangan, Mrs. Amy Heinz and her third grade class, for without their support my study would not be possible.

Finally, I wish to thank my cohorts for providing the initial reading and insight into the literature review. It was, without a doubt, very helpful.
TABLE OF CONTENTS

ABSTRACT...........................................................................................................iv
DEDICATION.........................................................................................................vi
ACKNOWLEDGEMENTS.....................................................................................vii
LIST OF FIGURES...............................................................................................xi
LIST OF TABLES....................................................................................................xii
CHAPTER I: INTRODUCTION..............................................................................1
CHAPTER II: LITERATURE REVIEW......................................................................4
    Assessment Efficiency and Reliability............................................................4
    The Utility of CBM and DIBELS Next.............................................................6
        DIBELS Next..............................................................................................6
    Defining Terms..............................................................................................9
    Legislation Initiatives....................................................................................11
    NCLB Reading Programs..............................................................................12
    Reauthorization of Elementary and Secondary Education Act of 2010...........13
    Decision Making Models for Reading Assessment........................................13
        Reading Assessment Models...................................................................13
Appendix C: Sample DIBELS Next Oral Reading Fluency Probe.........................43
Appendix D: Data Collection Tool........................................................................45
Appendix E: Random Assignment of DIBELS Next Passages.................................46
Appendix F: Scoring Rubric for DORF ADMINISTRATION.................................47
LIST OF FIGURES

Figure 1  Scatter Plot..................................................................................28
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>District Demographics</td>
<td>22</td>
</tr>
<tr>
<td>Table 2</td>
<td>Distribution of Scores</td>
<td>27</td>
</tr>
<tr>
<td>Table 3</td>
<td>Linear Relationship of Independent Variables</td>
<td>27</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

The increased need to screen, monitor, and assess student progress presents time management challenges to school personnel. This is evidenced in federal mandates and initiatives that require accountability through continuous progress monitoring and assessment responsibilities from the school’s paraprofessionals, teachers, and psychologists. Such is the case with the sweeping legislation of No Child Left Behind (NCLB) of 2001 and its effect on school testing and assessment programs (NCLB, 2001). Setting high standards for all students is the goal for which the selection of appropriate measurement tools is of primary concern (Stecker, Lembke, & Foegen, 2008).

Reading is critical to and is an overall predictor of academic success (Hosp & Fuchs, 2005; Stanovich, 2008). Central to student skill development is the use of reading assessment data for progress monitoring that will inform instructional planning. Regular and frequent screenings are critical to better learning outcomes for children (Goffreda, 2009). Stanovich (2008) suggests that reading should be assessed at an early age in order to intervene if necessary. Stanovich observed what he calls the “Matthew Effect” between proficient and struggling readers. The “Matthew Effect” is based on the biblical passage of Matthew 25:29 which states “that the rich will get richer and the poor will get
poorer”. This principle is apparent, he notes, in proficient and struggling readers. One will use existing reading skills to become more proficient or, because of a deficiency in skills, skills are used even less and the individual will fall further and further behind.

Psychometric instrumentation in reading that is efficient and dependable has led to widespread use of noncommercial and commercial curriculum-based measures (Griffiths, VanDerHeyden, Skokut, & Lilles, 2009; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). A commercial measure of early reading skills provides decision-making capability to countless educators in making strategic decisions for the reading needs of children. One such instrument, the Dynamic Indicator of Basic Early Literacy Skills (DIBELS) Next, Oral Reading Fluency (DORF) score is estimated by the number of words read correctly per minute (Good, Kaminski, & Moats, 2003). The DORF testing procedure is useful in exploring efficient time management procedures for teachers and other school personnel who are inundated with many testing responsibilities. The necessary and repetitive nature of assessing learning can take considerable time away from instruction. Because testing procedures are often vital to the learning-teaching process, research that focuses on effective yet efficient testing methods can be central in freeing up critical teaching time.

Another important reason for exploring efficient and reliable test measures is to validate the instruments’ utility in decision-making. In guiding instructional decision-making, Stecker, Lembke, and Foegen (2008) suggests a progress monitoring blueprint that includes five steps. These steps include: 1) select appropriate measurement materials, 2) evaluate technical features, 3) administer and score measures, 4) use data for goal setting, and 5) judge instructional effectiveness. Specifically, step two involves
examining the technical adequacy of the measurement tool. They suggest that measurement tools must be reliable, valid, and sensitive to change.

Thus, in examining the state of commercial and noncommercial measurement tools and in application of these tools to instructional practice, one possible research question arose: “Was the current three probe administration for which the median data is used and upon which many educational decisions are based, the most reliable and efficient method for collecting data?” An exploration of alternative methods is warranted in order to explore more practical alternatives in efficient educational decision-making.

The present study explored an alternative curriculum based measurement procedure and its reliability for practical applications in the educational setting. Twenty third grade oral reading passages were selected from the Dynamic Indicator of Basic Early Literacy Skills (DIBELS) system and used to assess reading progress of three third grade classes from different school districts in southwestern Ohio.

The standard procedure of using the one-minute median probe of a three probe administration was compared to the alternative procedure of using the first of three probes instead. The data was analyzed for alternate consistency. It was predicted that values for alternate reliability coefficients on single administration probes would be relatively reliable ($r \geq .70$) for participants. This study’s discussion included the innovation of progress monitoring tools, the utility of such, legal enactments, a decision making framework in reading for progress monitoring use with the K-12 school system and exploration of a more efficient and reliable solution.
CHAPTER II
LITERATURE REVIEW

Assessment Efficiency and Reliability. Assessment efficiency and reliability has been greatly influenced by the enactment of federal laws and the subsequent development of progress monitoring systems (Stecker, Fuchs, & Fuchs, 2008). Administrators, school psychologists, counselors, and teachers are increasingly held accountable for students’ performance. Thus, demand is high for suitable assessment and progress measures (Madelaine & Wheldall, 2004). Progress monitoring innovation has improved efficiency and dependability in teaching methods, accountability, and decision-making (Stecker et al., 2008). Educational decisions for reading are made at the federal, state, school district, and classroom levels. Federal and state decisions are made with assessment data for funding and effectiveness of educational systems. School districts are concerned with instructional programs, and teachers want to know how to gauge their teaching practices for maximum student learning (Stecker, Lembke, & Foegen, 2008). Measurement tools and procedures are central to the education process.

One measurement tool that satisfies both conditions of efficiency and reliability is the curriculum-based measurement (CBM) which was developed from the work of Stan Deno and Phyllis Mirkin in the late 1970’s and early 1980’s at the Minnesota Institute for
Research on Learning Disabilities (D. B. Marston, 1989). A CBM is a standardized assessment tool that is aligned with curriculum, is technically adequate, typically makes use of criterion-referenced measures, and is used over time to efficiently make low inference decisions for students (Hosp, Hosp, & Howell, 2007). A number of schools and districts in the United States and Canada use CBM as an alternative to other more cumbersome assessment methods in order to make appropriate instructional decisions (Hintze, Christ, & Methe, 2006; Hoffman, Jenkins, & Dunlap, 2009). However, it can be a time consuming endeavor to develop and maintain normative data and probe materials for ongoing assessment and monitoring. Busy teachers and educators are finding commercial CBM tools to be more practical. Specifically, in assessing oral reading fluency, there are several commercial CBM tools available. These tools include AIMSweb Standard Reading Assessment Passages (RAPs), Gray Oral Reading Test (GORT-4), the National Assessment of Educational Progress (NAEP) Fluency Scale, and Reading Fluency Monitor by Read Naturally (Hudson, Lane, & Pullen, 2005).

The commercial CBM instrument for measuring early literacy skills, which was explored in this study, is the Dynamic Indicator of Basic Early Literacy Skills (DIBELS Next) (Hudson et al., 2005). “Next” is part of the new version’s trade name indicating a progression from the 6th edition of DIBELS. DIBELS Next is a comprehensive assessment system that is validated through research with an estimated service base of over 1.7 million K-3 students. It is a set of procedures and short one-minute fluency measures that assess early reading skill development from kindergarten through sixth grade. DIBELS Next is the proprietary name given to the work of Roland Good and Ruth Kaminski and their associates at the University of Oregon (Hoffman et al., 2009). It
assesses the five big ideas in early literacy identified by the National Reading Panel Report (2000): Phonemic Awareness, Alphabetic Principle, Accuracy and Fluency, Vocabulary, and Comprehension.

The fluency performance measure used in both commercial and noncommercial CBM for reading, spelling, written expression, and mathematics computation developed as a result of a 1978 research-funded project awarded by Congress to Stanley L. Deno and others for designing and developing tools for teachers who needed to write and monitor Instructional Educational Plans (IEPs; Deno & Mirkin, 1978). The project’s goal was to identify key behaviors that would serve as simple, but general measures of academic performance. Reading fluency is correlated with key behaviors for academic success (D. B. Marston, 1989) and it serves as a screening tool for other reading skills (Stecker, Fuchs, & Fuchs, 2005). The ultimate goal of reading is to comprehend the text. The objectives in reaching comprehension include the same basic steps of fluency which are accuracy in word decoding, automaticity in recognizing words, and meaningful oral expression while reading (Rasinski, 2006).

The Utility of CBM and DIBELS Next

DIBELS Next. The DIBELS Next is a collection of reading probes which measure early literacy proficiency and identify at-risk students through short, criterion-referenced measures. Measurement for benchmarks is administered three times per year and each score value is determined by the student’s expected performance according to his or her academic grade level. Scores are identified for three risk categories: low-risk, some-risk, and at-risk (Goffreda, Diperna, & Pedersen, 2009). The low-risk category targets about 80 - 90 percent of the student body for proactive and antecedent
interventions. Antecedent interventions look at what in the environment can be addressed in order to prevent a problem from occurring (Gimpel Peacock, 2010). General education curriculum is targeted for these students. Students who fall between the twentieth and fortieth percentiles using local district norms, about 10 to 15 percent, should be considered to be in the some-risk category. Group interventions are targeted for these students. Finally, the at-risk category of students, about 5 to 10 percent of the population, is deemed to require high intensity, assessment-based, and individualized interventions.

One element of DIBELS Next is the DORF measure which is a standardized, individually administered test of accuracy and fluency with connected text (Good, Kaminski, & Moats, 2003). DORF is intended for most students from mid first grade through third grade. It is a set of passages and administration procedures for the purpose of identifying children who may need instructional support and for monitoring student progress. Third grade students may need intensive instructional support if they score below 70 in the spring of third grade.

In theory and best practices, teachers should usually administer the oral reading fluency measure once or twice weekly, or regularly at equal intervals throughout the school year (Deeney, 2010; Stecker et al., 2005). Conventional scoring protocol requires that words read correctly per minute (WRCM) are tallied and graphed. Words are supplied when the reader hesitates for three seconds, and are counted as errors. Self corrections made within three seconds are counted as correct. Also, mispronunciations, substitutions, omissions, and transpositions are counted as incorrect. Median scores are graphed along with trend lines and goal attainment (Stecker et al., 2005). However, some
researchers note that because of the popularity of and focus on the oral reading fluency metric; teachers might set student goals for achieving rates of speed only, a superficial demonstration of fluency (Deeney, 2010; Hasbrouck & Tindal, 2006; Samuels, 2007).

There is substantial support for the utility of DIBELS Next in making educational decisions in either a nomothetic or ideographic context (Christ, Johnson-Gros, & Hintze, 2005). Nomothetic applications are used to make norm-referenced decisions such as instructional groupings and screenings (Christ et al., 2005). It is here where relative interpretations are made in order to make norm-referenced decisions. Norm-referenced measures refer to the relative interpretation of test scores within a group (McKenna & Stahl, 2009). The ideographic utility in educational decisions uses the absolute interpretation of assessment outcomes to make criterion-referenced decisions. Criterion-referenced measures are used when there are specific benchmarks for which standards or proficiencies are to be met (McKenna & Stahl, 2009)
Defining Terms

Reading assessment parameters can be viewed and expressed both dimensionally, which refers to degrees or levels of reading, and categorically, which refers to basic, proficient, and advanced readers (Sattler & Hoge, 2006). The DIBELS NEXT/ORF assessment tool measures constructs related to early reading skills. These constructs are expressed both dimensionally with the use of percentile cut offs and expressed categorically with the use of progressive tiers that include: At Risk, Some Risk, and Low Risk (Hunley, 2010). Hosp (2007) suggests nine attributes that should be present in CBM methodology:

1. Alignment - what is being tested is congruent with what is being taught.

2. Technical adequacy - the methodology has been tested for validity and reliability. Validity is the degree to which a test measures what it claims to measure, and reliability is the consistency or dependability of the test's results. Early on, Marston (1989) summarized and established the reliability of CBM using psychometric data. Overall reliability calculated for the 6th edition of DIBELS Next ORF coefficients on the alternative-forms was .94 (Good, 2003), indicating a strong relationship between the testing instrument and its alternate form dependability between the probes.

3. Criterion-referenced - CBM typically makes use of criterion-referenced measures instead of norm-referenced measures. As mentioned earlier, norm-referenced
4. refers to a way of interpreting test scores by comparing the individual student’s results with that of other test takers. Criterion-referenced measures interpret results by looking at a pre-established criterion or benchmark (McKenna & Stahl, 2009). Criterion-referenced measures typically make use of criterion-referenced measures instead of norm-referenced measures. As mentioned earlier, norm-referenced refers to a way of interpreting test scores by comparing the individual student’s results with that of other test takers. Criterion-referenced measures interpret results by looking at a pre-established criterion or benchmark (McKenna & Stahl, 2009).

6. Standard procedures - following the same CBM administration and scoring rules.

7. Performance sampling - procedures that employ direct, low-inference measures (conjecture about meaning is kept to a minimum) where correct/incorrect responses are based on clearly defined tasks.

8. Standardized rules - established based on performance criteria through sampling or experimental procedures. These rules indicate what it means when students score at different levels of performance or different rates of progress over time.

9. Repeated measurement - CBM emphasizes repeated measurement over time which means that CBM administrations can be used for progress monitoring to determine the rate of learning growth.

10. Efficient - CBM is efficient in several ways: it is efficient in implementation because people can be trained in a short period of time to give the measure quickly; it is efficient in communication because behavior and performance data
can be interpreted from raw scores. Also, it can be administered to individual students in relatively brief periods of time.

Accessible - it is easily summarized and graphed thus increasing accessibility to teachers and students. Also, the data can be easily interpreted when graphed.

**Legislation Initiatives**

Reading content area measures have been addressed by both federal legislative mandates and initiatives, and continue to be a leading topic (Hosp et al., 2007). Reading fluency poses a wide range of questions and variables which need to be conceptualized and answered (Rasinski, 2005).

One such federal legislative reform is No Child Left Behind Legislation (NCLB) of 2001. NCLB is a national education reform policy which requires school accountability through challenging content standards for mathematics, reading or language arts, and science (NCLB, 2001). All curriculum goals should be based on measurable achievement standards (Jacob & Hartshorne, 2007). Coupled with this directive is the need to measure these content areas on a yearly basis with measures that meet national technical standards for both validity and reliability. These measurement outcomes can and do culminate in high and low stakes decision making. In high and low stakes decision making, educators must be able to distinguish between the benefits of ORF testing, the misuse of it, and how the testing will be used. The pressures have mounted upon schools to provide definitive answers for what ails our nation’s students and, and therefore, suggest a need to examine decision-making practices for curriculum, instruction, and assessment (Trybus, 2007).
The federal mandates of NCLB and the resulting reauthorization of the Individuals with Disabilities Education Act (IDEA 2004) are clear examples of some of the most wide sweeping educational reform legislation in United States history (Connelly, He, & Phillion, 2008). The initiatives of No Child Left Behind (NCLB), the National Reading Panel recommendations, and IDEA all impact educational outcome standards for the work of schools, teachers, and students. An initiative through the National Institute of Child Health and Human Development (Cunningham, 2001) included a similar directive to teachers advocating for regular fluency assessment. One broad mandate and implication of the NCLB Act of 2001 is the term "scientifically-based research." NCLB sets mandates for states to employ a scientific research-based approach when identifying students with learning disabilities, and in making curriculum, monitoring, and diagnostic decisions (Trybus, 2007). This means that research must use empirical methods that are based on an experimental design, rigorous data analysis to test the hypothesis, measurements that produce valid data across evaluators and observers, and be accepted by a peer-reviewed journal or a panel of experts (Trybus, 2007). This approach is conceptualized in the Response to Intervention (RTI) initiative for statewide reading and mathematics testing, which are implemented annually in grades three through eight (Jacob & Hartshorne, 2007).

NCLB Reading Programs

There are six reading programs emphasized in NCLB which include Reading First, Early Reading First, William F. Goodling Even Start Family Literacy Programs, Improving Literacy through School Libraries, Reading is Fundamental—Inexpensive Book Distribution Program, and Striving Readers (United States Congress Senate
Committee Report on Health, Education, Labor, and Pensions, 2002). All of these programs include research based interventions and assessment. Furthermore, the Reading First program explicitly states that classroom-based screenings using instructional and diagnostic reading assessments are required (Moss, Abt Associates, & United States, 2008).

Reauthorization of Elementary and Secondary Education Act of 2010

The aim of the reauthorization of the Elementary and Secondary Education Act released in March 2010 is to raise accountability standards from the current state and focus its scope to “college and career ready standards” for all students (United States Congress House Committee Report on Education and Labor, 2010). Formula funds will be available only to states implementing these objectives by the year 2015. Comprehensive, pre-K-12 literacy programs are to be planned and implemented for improving student achievement. Other areas for improvement include science, technology, engineering, and mathematics (STEM). The education reform explicitly states that priority funds will be given to programs that demonstrate evidence of effectiveness and efficiency in the use of resources (United States Congress House Committee Report on Education and Labor, 2010).

Decision Making Models for Reading Assessment

**Reading Assessment Models.** McKenna and Stahl (2009) identify four types of models for reading assessment: the Deficit Model, the Contextual Model, the Stage Models, and the Cognitive Model. These model types provide different perspectives and assumptions for the source of the reading problem. The deficit model is sometimes referred to as the medical model and it assumes that reading problems within a child need
a diagnosis for eventual remediation. On the other hand, the contextual model assumes that reading difficulty arises from environmental situations that, if addressed, could solve the problem. However, there is little evidence for the Deficit or Contextual model of reading assessment as an explanation for reading problems (McKenna & Stahl, 2009). In addressing the remaining two models, which are both empirically verified, the question becomes whether either or both of these models can accommodate DIBELS Next ORF data in the larger context of educational decision-making. The stage models were developed in an effort to explain reading progression and its potential problems. These models, including the Chall model, and the Spear-Swerling and Sternberg’s model, attempt to explain overall reading or the specific analysis of reading difficulties (McKenna & Stahl, 2009).

Finally, there are three separate components suggested by the Cognitive approach that include, 1) automatic recognition of the words in the text, 2) comprehension of language in the text, and 3) the ability to strategize in order to obtain one’s purpose in reading the text (McKenna & Stahl, 2009). The difference between stage models and the cognitive model is that the former focuses on the process of reading whereas the latter focuses on the components of reading acquisition or breakdown. As mentioned in prior discussion, DIBELS Next uses fluency measures as an indicator of broad skill areas. This provides a general heuristic of reading that includes a number of fluency measures. These measures target initial word sounds (ISF), sounds within a word or phonemic segmentation (PSF), letter-sound knowledge or nonsense words (NWF), grade level text reading or oral reading (DIBELS Next ORF), text understanding or retell, and vocabulary or word use (WUF). DIBELS Next could provide assessment for all of the emergent 
reading stages that are identified by the Spear-Swerling and Sternberg model (1996) and
the Cognitive model (McKenna). However, the advanced level of phonological
awareness such as strategic reading, the final stage of the Spear-Swerling and Sternberg
model, would not be accommodated by the DIBELS Next measure, nor does DIBELS
Next purport to assess these advanced skills.

Response to Intervention (RTI). DIBELS Next also aligns itself with the three
tiered Response to Intervention (RTI) Model. Both DIBELS Next and RTI present three
levels of assessment for intervention. The DIBELS Next assessment categories, as
previously mentioned, include deficit, emergent, and established, or at risk, some risk,
and low risk. RTI assessment categories include universal core instruction, strategic, and
case study for intensive intervention with possible referral for special education. RTI
provides a broad framework in which reading can be assessed in both relative and
absolute terms (Hunley, 2010; Kaminski et al., 2008) It is a process for evaluating
whether students respond to evidence-based instruction (Stecker et al., 2008) and for
identifying various levels of instructional need (Lembke, McMaster, & Stecker, 2010).

Within the RTI model is a framework known as the School-Wide Model for
Reading Intervention (SWM-RI; Shinn, 2002). This assessment system primarily uses
DIBELS Next and is used in kindergarten through third grade. Usually, implementation
begins with a school-based team which determines what the current reading practices are
and whether a new plan should be put in place (Shinn, 2002). SWM-RI uses a three
tiered assessment for intervention approach where 80% of students in each grade would
be targeted for general core reading curriculum, 10-15% of students would receive
strategic reading support, and 5-10% of students would receive intensive support.
However, David Chard cautions educators that some students, in spite of the multilevel system of prevention and intervention, will still struggle with reading difficulties (Chard et al., 2008). This is why, he says, it is important to understand student learning.

**Problem-solving Model.** A problem-solving process perspective that builds on local reading norms can assist decision makers in identifying problems and in determining whether additional resources are needed, such as those available through special education. The Problem-Solving model rests on three basic assumptions: 1) a problem is defined as a discrepancy between what is expected and what actually occurs; 2) there is a subset of students whose differences are so significant that a modification to the general curriculum is warranted; and 3) educators must generate many plans of action or hypotheses before attempting problem solutions (D. B. Marston, 1989; Shinn, 2002). The problem-solving framework seeks to answer four basic questions through the application of science methods: 1) Problem Identification – Is there a problem and what is it? 2) Problem Analysis – Why is the problem happening? 3) Treatment Design – What can be done about the problem? 4) Treatment Evaluation - Did the intervention work (Caplan, 1970)?

Federal and state legislation, and potential funding sources, have each imposed new and stringent timelines on decision-making for student placement and progress (Stecker et al., 2008). Decision-making that is based on the empirically-based models such as the stage and cognitive models, provide an efficient and dependable way to measure, diagnose, and monitor students’ performance in reading. When this approach is placed within the larger context of RTI, SWM-RI, and Problem-Solving, DIBELS Next ORF provides a practical measurement tool within the current school environment.
Exploring the DIBELS Next ORF instrument for efficiency and reliability is a front line attempt at effective educational decision-making (Green & Salkind, 2005).

**Exploring an Efficient and Reliable Solution**

Meaningful and useful educational decisions must be based on measurement tools that are dependable and consistent (Sattler & Hoge, 2006). The reliability characteristics of dependability and consistency can be viewed in one of four ways. A measure may be consistent within itself which demonstrates internal reliability; a measure can be consistent over time which demonstrates test-re-test reliability; a measure can be consistent with a parallel form which demonstrates alternate-form reliability; or a measure can be consistent when used by another rater/observer which demonstrates interrater or interobserver reliability (Sattler & Hoge, 2006). In theory, reliability of measurement is determined by the degree of change in score administrations. The more random and less systematic the change of scores during repeated administrations, the less reliable the measurement tool.

Sattler writes:

According to classical psychometric theory, a test score is composed of two components: a true score and an error score. True score refers to the measurement process, and error score is any factor that contributes to inconsistencies from test to test. Both the true and error scores combine to create the observed score that is seen in testing. The theory assumes that 1) the examinee possesses stable traits; 2) errors are random; and 3) the observed test score results from the addition of the true score and the error score (Sattler & Hoge, 2006, p.46).
In the current study, true score is evidenced by the consistency in rate of oral reading fluency improvement over time, and error is explored in relation to alternate forms of DIBELS Next probes.

Educational decision making that is based on measurement outcomes is influenced by dimensions, characteristics, and arrangement of stimuli both within and across CBM (Christ & Vining, 2006). Several studies have documented the need for determining efficient measurement methods and its decision-making utility through the arrangement of test characteristics. One such study by Christ and Hintze (2005) demonstrated the need for altering assessment durations in assessing multi-skill computational math fluency as a way of making better educational decisions based on curriculum-based outcome measures (Christ et al., 2005). The results suggest that adequate confidence was attained to make low stake decisions for 1-minute administrations and adequate confidence was attained to make high stake decisions by using relative score interpretations after a single 4-minute administration.

In another study in which Ardoin and Christ (2008) examined procedures for conducting triannual universal reading screenings, CBM slope estimates were calculated using single and median DIBELS Next scores across universal same and different passage sets. Results suggested that same probe sets be used across universal screenings and shorter time frames such as semester as opposed to annual rates of growth be used for evaluation purposes.

A later study (Ardoin & Christ, 2009) evaluated and estimated the magnitudes of standard error across an experimental passage set, and two commercial passage sets of AIMSweb, and DIBELS. Results suggested that estimates were smallest in magnitude
for the experimental passage set, followed by AIMSweb and then DIBELS. Standard errors are estimated to stem from variations in reading passage difficulty. These reliability studies explored alternate ways of administering assessment and the implications for making educational decisions.

**The Present Research Study**

Very little research has been done on median versus single probe administrations. More studies are needed that explore and replicate median versus first probe utility. This study explored the alternate form characteristic of first probe administration in comparison to the median probe on several reading forms of the DIBELS Next ORF measure. The primary focus explored an alternative curriculum based measurement (CBM) procedure and its reliability for practical applications in the educational setting. The standard procedure of using the median probe score of a three probe administration was compared to the alternative procedure of using the first of three probes administered. Results are discussed in terms of first and median probe administration consistency and its implication for efficiency in decision making.
CHAPTER III

METHOD

Research Question and Hypothesis

The research question that was investigated in the current study was: To what extent do the median raw scores correlate with the scores of the first probe administered? It was hypothesized that values for alternate form reliability coefficients on single administration probes would be relatively reliable (r ≥ .70). This hypothesis would be accepted if there was a relatively reliable level (r ≥ .70) of consistency on the alternate form coefficients for probe administrations during weeks one through ten using all participants. Conversely, the null hypothesis would be accepted if values for reliability coefficients were not relatively reliable (r ≤ .70) for administrations during weeks one through ten using all participants.

Research Design

This study used a quasi-experimental design in order to confirm or disconfirm the hypothesis that the DIBELS Next ORF measure is relatively reliable with a single administration probe. A correlational analysis was conducted between two raw scores: single (first) administration probe score, and the median administration probe score.
Participants and Setting

Each school was located in a different school district in southwestern Ohio, which were identified as one rural and two urban/suburban. One third grade class was targeted at each school. Total participants from all three third grade classes included 62 students. The three participating investigators/school psychologist interns were responsible for obtaining permission for student participation from the school principal (see Appendix A), the classroom teacher (see Appendix B), and oral assent from the students themselves. The three school districts varied in size and demographics. Information regarding each participating school district can be viewed in Table 1. Demographic composition of the sample included the following: Third graders, 100%; Asian/Pacific Islander, 16%; Black-nonHispanic, 7%; White-nonHispanic, 74%; Other Ethnic Origin, 3%; Males, 47%; and Free Lunch, 35%.

Instrument

Materials. Materials included a student copy of the DIBELS Next ORF passages, an examiner copy of the passage (Appendix C), a data collection sheet (Appendix D) a clipboard, a timer, and a colored scoring pen (Good, Kaminski, & Moats, 2003). The DIBELS Next ORF passages were obtained online.

The previous DIBELS DORF, 6th edition measure has technical adequacy (Good, Kaminski, & Sopris West Inc, 2003; D. B. Marston, 1989). The degree of test reliability can be expressed in a coefficient from 0 to 1. The closer the reliability coefficient is to 1, the more reliable the procedure. There are several procedures to determine whether a test is consistent or reliable. Alternate Form reliability is a measurement that indicates how consistent two forms of a test are, Test-retest
Table 1
District Demographics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>District A</th>
<th>District B</th>
<th>District C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Daily Student Enrollment</td>
<td>2419</td>
<td>7571</td>
<td>5338</td>
</tr>
<tr>
<td>Black non-Hispanic</td>
<td>0.4%</td>
<td>3.0%</td>
<td>6.9%</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>---</td>
<td>0.2%</td>
<td>---</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>---</td>
<td>6.5%</td>
<td>2.4%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.5%</td>
<td>2.6%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Multi-Racial</td>
<td>3.0%</td>
<td>3.7%</td>
<td>3.8%</td>
</tr>
<tr>
<td>White non-Hispanic</td>
<td>94.7%</td>
<td>84.0%</td>
<td>84.8%</td>
</tr>
<tr>
<td>Economically Disadvantaged</td>
<td>44.4%</td>
<td>13.2%</td>
<td>36.0%</td>
</tr>
<tr>
<td>Limited English</td>
<td>---</td>
<td>2.4%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Students with Disabilities</td>
<td>16.4%</td>
<td>13.3%</td>
<td>14.3%</td>
</tr>
</tbody>
</table>

Note. --- = Not calculated/Not displayed when there are fewer than 10 in the group. Information was taken from most recently reported Local Report Card data at: [http://ilrc.ode.state.oh.us/default.asp](http://ilrc.ode.state.oh.us/default.asp)
procedures measure the performance of a test taker on the same test taken on different occasions; Internal Consistency measurement that compares parts within a test for consistency; and Inter-rater reliability where test outcomes are compared from more than one rater for consistency. Test-retest reliabilities for elementary students range from .92 to .97; and alternate-form reliability of different reading passages drawn from the same level ranged from .89 to .94. For the DIBELS, studies confirm good reliability for the DORF measure. Reliability coefficients on DORF were as follows: Third grade Alternate Form Reliability when two forms or single passages of the tests were compared for WRCM was .89; Third grade Test-Retest Reliability where test performance was compared for the same test on different occasions for WRCM on DORF Triad was .93 (triad refers to 3 levels of reading difficulty determined by the Spache Readability formula randomly used throughout the grade level reading passages); and third grade Inter-Rater Reliability where tests outcomes were compared from more than one rater for consistency of WRCM and using the DORF Triad was .99 (Powell-Smith, Good, & Atkins, 2010).

Procedures

Administration. Three 1-minute DIBELS Next Oral Reading Fluency probes were administered individually to three classes, totaling 62 students once a week for 10 weeks, beginning the 4th week of the academic year. Students were administered three DORF passages in a single sitting. Scores were based on the number of words read aloud correctly within the 1-minute time period. Individual administration time for all three probes was approximately 6 minutes. The 20 different reading passage probes were administered in random order to control for practice effects (see Appendix E). Some
probes were administered more than once. Griffiths et al. (2009) found that CBM probes can be used for benchmarking and then again for progress monitoring without a reread effect, even when the same probe is used more than once. Scores for all three probes were recorded along with the basic descriptive data obtained from school records for each student including: gender, ethnicity, date of birth, and eligibility for free or reduced lunch which represented socio-economic status. An example of a DIBELS Next probe is provided in Appendix C. Standard administration procedures (Good, Kaminski, & Moats, 2003) were followed by graduate-level interns in school psychology for all students using a documented integrity checklist which includes the following:

1. A student copy will be placed in front of the student and the examiner copy will be placed on a clipboard and positioned so that the student cannot see what is recorded.

2. The administrator will give the following specific directions: “Please read this (point) out loud. If you get stuck, I will tell you the word so you can keep reading. When I say, ‘stop’ I may ask you to tell me about what you read, so do your best reading. Start here (point to the first word of the passage).”

3. A stopwatch will be used to begin timing when the student says the first word of the passage (not including title).

4. If the student fails to say the first word after three seconds, the administrator will tell the student the word, mark as incorrect, and begin timing when the student reads a word independently.

5. If the student does not provide a word within 3 seconds, the test administrator will say the word and mark the word as incorrect.
6. Errors are marked on the examiner copy of the probe by putting a slash (/) over words read incorrectly.

7. At the end of one minute, place a bracket ([]) after the last word provided by the student.

8. Stop and reset the stopwatch, and say “stop” (remove the passage).

**Scoring.** Reading passages were scored and recorded using scoring booklets that were transferred to a scoring rubric after administration (Good, Kaminski, & Moats, 2003).
A parametric statistical analysis was conducted on the blinded data using SPSS. Students were identified by first name or initials along with demographic data only. Data were analyzed using Pearson correlation coefficients for two correlates, single and median probe data in order to obtain an alternate-form coefficient and expressed as an “r” value ranging from 0 to 1. This coefficient indicated the degree to which low or high scores on single probes compared to low or high scores on the median probes (Green & Salkind, 2005).

The analysis was completed on progress monitoring data collected from 3 probe administrations over 10 weeks for all 62 students. The results of the correlational analysis presented in Tables 2 and 3 show that the correlation or linear relationship between the independent variables was statistically significant at the .01 level. Table 2 shows the mean difference between all probe 1 scores to that of median probe scores.
Table 2

Distribution of Scores

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1</td>
<td>105.02</td>
<td>37.485</td>
<td>539</td>
</tr>
<tr>
<td>Median Probe</td>
<td>106.20</td>
<td>38.161</td>
<td>539</td>
</tr>
</tbody>
</table>

Table 3

Linear Relationship of Independent Variables

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Statistical Analysis</th>
<th>Probe 1</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.958</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>539</td>
<td>539</td>
</tr>
<tr>
<td>Median</td>
<td>Pearson Correlation</td>
<td>.958</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>539</td>
<td>539</td>
</tr>
</tbody>
</table>

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

The results are provided as a scatter plot in Figure 1. The correlation of single probe scores was positive and highly significant with median score probes, r (537) = .96, p < .01. Single probe administration accounts for 92% of the variability in median probe
administration. In general, the results suggest that low and high scores are associated equally on the single probe administration as they are on the median probe administration.

Figure 1 Scatter Plot
CHAPTER V
DISCUSSION

**Review of Purpose and Major Findings.** Research that examines DIBELS Next ORF’s alternate form reliability can begin the process for technical adequacy in CBM procedure, its application, and its interpretation. If alternate form consistency of the first probe and median probe is low ($r < .70$), then the test’s decision making utility may be limited depending on the reliability estimate of the type of decision making associated with the test results. The use of median scores with three probes instead of single administrations may be prudent when reliability is low ($r < .70$) for some decision making contexts.

The overall purpose of this study was to provide first line data for the exploration of CBM probe administration efficiency. CBM is part of the academic assessment process and is a critical part of educational decision-making. Efficient and reliable CBM procedures will allow educators to have decision-making confidence and to devote more classroom time to teaching.

The major finding of this study was that high and low scores were equally associated on single and median probe administration. This means that the DIBELS Next ORF reading passage probes when compared in alternate form for first and median
probes were reliable and were administered with little error. This evidence suggests that the standard and best practice of median probe administration developed in the late twentieth century as a result of research by Stanley Deno and Phyllis Mirkin, may be outdated, time consuming, and unnecessary in making some decisions. Instructional decisions for screening, progress monitoring, diagnostics, and outcome made for individuals, small group, class, or school-wide initiatives may be safely determined with one probe instead of three, once reliability on a measure is established.

This study’s data could be interpreted as an anomaly because all of the reliability factors which were in place such as heterogeneity of sample, and probes given within a six-minute time span, would not always be necessarily present conditions in a real world scenario. Thus, effort and attention should be made to ensure that reliability conditions are established for a measure.

**Limitations.** This study was limited in the convenience sampling procedure. Selecting participatory schools posed some restrictions in the randomized sampling assumption. Testing procedures that are numerous and repetitive require school personnel who are willing and have resources of time, personnel, and space for such an endeavor. Some schools do not use the DIBELS Next measure and were consequently unwilling to participate in the study. Also, another limitation might include the training and experience that each graduate student had prior to implementing the DIBELS Next data collection procedure. The three participating graduate students had varying degrees of training and experience with administering DIBELS Next procedures.
**Future Research.** Future research should focus on identifying practical ways of including reliability factors in the educational setting. For some tests, longer probes may be warranted because the longer the test duration the more reliable it will be (Christ et al., 2005). In general, future research should explore what are the necessary and sufficient reliability conditions in the educational setting.

One factor for increasing reliability of CBM is to decrease measurement error. Neddenriep, Poncy, and Skinner (2011) suggest four factors when using CBM data to support educational decisions:

1) Characteristics of the selected probe set, which refers to the consistency of probe difficulty; Error is reduced when the probe set is tightly controlled such as in DIBELS or AIMSWeb;

2) The type of decisions being made refers to whether the decision is relative (comparing students to one another) or absolute, which refers to determining individually a student’s performance and/or progress. Error is reduced in relative decisions because all students are presumed to receive the same amount of treatment error, so student ranking would not be affected;

3) The standardization of administration conditions, which refers to consistency in administration. Consistent administration procedure produce reliable data; and

4) The number of behavioral samples taken, which is at the center of this study. Conventional wisdom supports the notion that more probes are useful and reduce error with absolute decision making. Again, relative decisions are not affected by the number of probes.
Conclusion. In summary, data based decision making in the schools has been shaped through current legislative initiatives to increase accountability. This increase in assessment activity has led to the use of CBM as an assessment tool in teaching, progress monitoring, and establishing benchmark goals. Its utility is encouraged in order to make effective educational decisions for children. Research that begins the discourse of the necessary and sufficient conditions for aligning CBM as an assessment tool with efficiency and reliability is a necessary step in making effective educational decisions. This study explored a less time consuming data collection procedure that could have tremendous benefit for educators who are strapped with resource constraints of personnel and time.
REFERENCES


University of Dayton DIBELS NEXT Research Project
I, ____________________________, give my permission for research to be done in
_____________’s classroom in my school, __________________________. I
understand the following:

- The entire class will be participating in the research
- The research will be conducted by a graduate student under the supervision of a licensed
  school psychologist and faculty of the University of Dayton
- The purpose of the research is to better understand a child’s reading progress
- All children will benefit from weekly monitoring of reading progress
- Student records will be accessed for demographic information (i.e., gender, ethnicity, free
  and reduced lunch, date of birth)
- All records will be kept confidential using a lock box and no names will be used
• Participation is voluntary and there will be no penalty or loss of benefits because of refusal to participate and participation can be discontinued at any time.

• The contact for this project is:
  
  o Debbie Gillespie, M.S: (937) 376-7729  lilsanct7@woh.rr.com  or
  
  o Sawyer Hunley, PhD: (937) 668-4005, shunley1@notes.udayton.edu

Principal Signature: _____________________________   Date: _______________
APPENDIX B

TEACHER PERMISSION FORM

University of Dayton DIBELS NEXT Research Project
I, ____________________________, give my permission for research to be done in my classroom. I understand the following:

• The entire class will be participating in the research
• The research will be conducted by a graduate student under the supervision of a licensed school psychologist and faculty of the University of Dayton
• The purpose of the research is to better understand a child’s reading progress
• All children will benefit from weekly monitoring of reading progress
• Student records will be accessed for demographic information (i.e., gender, ethnicity, free and reduced lunch, date of birth)
• All records will be kept confidential using a lock box and no names will be used
• Participation is voluntary and there will be no penalty or loss of benefits because of refusal to participate and participation can be discontinued at any time
The contact for this project is:

- Debbie Gillespie, M.S: (937) 376-7729  liisancet7@woh.rr.com or
- Sawyer Hunley, PhD: (937) 668-4005, shunley1@notes.udayton.edu

Teacher Signature: _______________________________ Date: _______________
A New Ball Game

On the first day of school, Roy’s teacher asked him to write a letter about himself. Roy was glad to have the chance to talk about his life in Africa. Roy had been born in the United States, but his family had lived in a small town in Africa for three years. Now his family had moved back to the United States.

Roy’s stepmom was a doctor. She worked in a clinic, where she treated sick people and gave immunizations. His dad taught music at the school Roy and his brother attended. Roy and his friends played sports together and practiced playing the instruments his dad taught them. Football was his favorite sport and there always seemed to be a game going after school.

In his letter, Roy wrote about his life and that he missed playing football the most. He had seen American football and it was a very different game. It was hard to figure out why they were chasing each other and when to cheer.

He handed the letter to his teacher the next morning. That afternoon as he was leaving, his teacher called him over. Another boy was standing next to her. “Roy, this is Spencer,” his teacher said. “He’s going to introduce you to the soccer team. I think you’ll
enjoy it.” Spencer smiled at Roy and led him outside to the field, where a group of kids were playing. “The teacher said you call this football in Africa, but here it’s called soccer,” said Spencer. Roy looked around.

Additional DIBELS NEXT passages can be obtained at

https://DIBELSNEXT.uoregon.edu/measures/index.php
APPENDIX D

DATA COLLECTION TOOL

Student identification: __________

For use with every student assessed using DIBELS Next.

<table>
<thead>
<tr>
<th>Administration Dates</th>
<th>ORF score 1:</th>
<th>ORF score 2:</th>
<th>ORF score 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ethnicity: White /Black/American Indian/ Eskimo/ Aleut/Asian/ Pacific Islander or Other

Free/reduced lunch? Yes No Gender: Male/Female

Date of Birth:
### APPENDIX E

**RANDOM ASSIGNMENT OF DIBLES NEXT PASSAGES**

<table>
<thead>
<tr>
<th>Week</th>
<th>District A</th>
<th>District B</th>
<th>District C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5   17  4</td>
<td>8   7  18</td>
<td>7   9  8</td>
</tr>
<tr>
<td>2</td>
<td>19  16  18</td>
<td>11  9   1</td>
<td>1   13  20</td>
</tr>
<tr>
<td>3</td>
<td>6   20  14</td>
<td>2   16  19</td>
<td>10  5  15</td>
</tr>
<tr>
<td>4</td>
<td>10  11  3</td>
<td>12  4   17</td>
<td>11  19  14</td>
</tr>
<tr>
<td>5</td>
<td>12  8   7</td>
<td>5   13  6</td>
<td>3   17  12</td>
</tr>
<tr>
<td>6</td>
<td>9   1   13</td>
<td>14  20  3</td>
<td>2   4   18</td>
</tr>
<tr>
<td>7</td>
<td>15  2   11</td>
<td>15  10  5</td>
<td>16  6  10</td>
</tr>
<tr>
<td>8</td>
<td>1   10  6</td>
<td>6   9   2</td>
<td>7   14  6</td>
</tr>
<tr>
<td>9</td>
<td>14  4   19</td>
<td>11  18  1</td>
<td>15  11  3</td>
</tr>
<tr>
<td>10</td>
<td>7   12  2</td>
<td>16  4   7</td>
<td>16  17  8</td>
</tr>
</tbody>
</table>
APPENDIX F
SCORING RUBRIC FOR DORF ADMINISTRATION

If student does not read any words correctly in the first row, test will be discontinued, and scored a zero on the front cover.

Otherwise the total number of words read correctly will be scored on the bottom of the scoring sheet.

If the student reads fewer than 10 words correctly on the first passage, the score will be recorded with no further administration of passages 2 and 3.

All 3 scores will be recorded if 3 passages are read.

If a word is mispronounced, it is scored as an error.

Numerals that are read correctly in the context of the sentence will be counted as correct.

Hyphenated words that can stand alone will be counted as two words and those words that cannot stand alone, will be counted as one.

If a student hesitates or struggles with a word for 3 seconds, the correct word will be given and the word will be scored as incorrect, and the student will be prompted to continue to the next word.

If the student self corrects, “SC” will be marked above the word and scored as correct.

Words that are repeated are ignored and scored as correct.

The student will not be penalized for articulation, second language and dialect pronunciation patterns.

Inserted words will be ignored and not counted as errors.

Abbreviations will be read in the way that they are normally pronounced in conversation.

All words that are read correctly but in the wrong order are scored as incorrect.

Omitted words are scored as incorrect.

Numerals that are read correctly in the context of the sentence will be counted as correct.

Hyphenated words that can stand alone will be counted as two words and those words that cannot stand alone, will be counted as one.

If a student hesitates or struggles with a word for 3 seconds, the correct word will be given and the word will be scored as incorrect, and the student will be prompted to continue to the next word.

If the student self corrects, “SC” will be marked above the word and scored as correct.

Abbreviations will be read in the way that they are normally pronounced in conversation.

All words that are read correctly but in the wrong order are scored as incorrect.

Omitted words are scored as incorrect.

If the student does not read any words correctly in the first row, test will be discontinued, and scored a zero on the front cover.

Otherwise the total number of words read correctly will be scored on the bottom of the scoring sheet.

If the student reads fewer than 10 words correctly on the first passage, the score will be recorded with no further administration of passages 2 and 3.

All 3 scores will be recorded if 3 passages are read.