1-2011

Stability of Popular R-CBM Progress Monitoring Tools: Dibels® Next and Aimsweb®

Christine Russell
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, and the Special Education and Teaching Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/458

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
STABILITY OF POPULAR R-CBM PROGRESS MONITORING TOOLS: DIBELS® NEXT AND AIMSWEB®

by

Christine Russell

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Special Education and Literacy Studies
Advisor: Kristal Ehrhardt, Ph.D.

Western Michigan University
Kalamazoo, Michigan
August 2011
Currently there is no agreed-upon method for determining the difficulty level, referred to as the readability level, of Reading Curriculum-Based Measurement (R-CBM) passages. A key tenant of R-CBM is that the passages across each grade level are equivalent in difficulty level and therefore can be used to monitor student academic improvement. The primary objective in this study was to evaluate the homogeneity of oral reading fluency progress monitoring passages of two popular passage sets that are used frequently in schools. The purpose of this research was to examine the stability of each R-CBM progress monitoring passage set as well as determine whether there is any benefit to organizing the progress monitoring passages into triad sets for interpretation. The results indicated even with the most current methods of equating progress monitoring passages, error related to passage difficulty continues to persist. It is clear that using strong tactics such as a well developed readability formulas, as well as field testing passages, leads to a better equated passage set. In addition, analyzing progress once there has been three assessments given across time, rather than after each individual progress monitoring session, leads to considerably better information regarding student reading growth with reduced error related to passage difficulty level.
ACKNOWLEDGMENTS

I would like to begin by thanking those who encouraged me in this endeavor. The process of completing a dissertation has been a challenging experience, but also has taught me an immense amount about scholarly research, self-discipline, and perseverance through barriers and obstacles.

Special thanks go to my doctoral committee. Thank you, Dr. Whitten, Dr. Morgan, and Dr. McGlinchey, for your willingness to give time to this project. Dr. Ehrhardt, all along the way through my education in school psychology and special education you have been supportive and have opened many doors and opportunities that have led me to where I am today. There have been times when you have thrust me into situations where I was unsure of how I would handle them, and these resulted in increased confidence and skills in areas such as instructing and writing.

Without the enthusiasm of the principals in Parchment Community Schools, Sarah Johnson and Marcy Patterson, and the flexibility of the teachers in the classrooms, this research would not have been possible. They made an experience that could have been difficult to manage run so smoothly without one complaint when timelines went longer or small barriers presented themselves. The students in Parchment are fortunate to have such wonderful leadership and staff. I am privileged to have gotten to work with all of you.

I was lucky enough to have an amazing group of research assistants assemble and give their time to this project. Andrea Wood, a particular thank you to you for your
positive attitude and amazing organizational skills. I appreciate all that you did to keep
the data collection moving. Additional individuals who gave their time to assist with data
collection—Rashell Bowerman, Cashmier Broten, Jennifer Meyer, and Maureen Turner,
thank you for your time, work, professionalism, and encouragement.

I would also like to thank Jason, my supportive husband, who endured hours upon
hours of losing his wife to time spent holed up working on this dissertation. I could not
have asked for deeper support. This would never have been possible without him there to
take on parenting when needed, and lend an understanding ear when necessary. I also
thank my sister Debbra for all of her encouragement through the last few years, much
needed prayers, and her sincere joy for me when small and large milestones were met.
Lastly, I thank Kaylin for being a daily inspiration to me, and Nolan and Cole, the baby
boy twins for cooperating enough that I was able to get so far in this project prior to their
arrival. I am blessed upon blessed.

Christine Russell
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... ii

LIST OF TABLES ................................................................................................................ vii

LIST OF FIGURES ................................................................................................................. viii

CHAPTER

I.  REVIEW OF THE LITERATURE ......................................................................................... 1
   Introduction to CBM ........................................................................................................ 1
   Introduction to R-CBM .................................................................................................... 2
   Practical Uses of R-CBM .............................................................................................. 4
   R-CBM in the Response to Intervention (RtI) Model .................................................... 5
   Research in R-CBM ........................................................................................................ 8
   R-CBM Readability ....................................................................................................... 10
   R-CBM Slope ................................................................................................................ 18
   DIBELS Version 6.0 and DIBELS Next Passage Development ................................. 24
   AIMSweb Passage Development ................................................................................. 26
   Research Questions .................................................................................................... 28

II.  METHOD ........................................................................................................................... 29
   Procedure .................................................................................................................... 30
   Participants and Setting ............................................................................................... 30
   Consent and Assent Procedures .................................................................................. 33
Table of Contents—Continued

CHAPTER

Materials ................................................................................................................. 34
R-CBM Probes........................................................................................................ 34
Training Procedures ............................................................................................... 35
Interrater Reliability ................................................................................................. 37
Procedural Integrity ................................................................................................. 37
Dependent Measures ............................................................................................... 38
    Oral Reading Fluency ......................................................................................... 38
Risks and Costs ........................................................................................................ 38

III. RESULTS ........................................................................................................... 40
    Overview .............................................................................................................. 40
    Self-Assessments, Procedural Checklists, and Interrater Reliability ............ 41
    Variance Within and Between Passage Sets .................................................... 42

IV. DISCUSSION ....................................................................................................... 47
    Analyzing Data in Triads .................................................................................... 48
    Effect of Readability Formulas on the Interpretation of R-CBM Data .......... 50
    Differences in the Grand Means of DIBELS Next and AIMSweb ............... 51
    Limitations ........................................................................................................... 51
    Conclusion ............................................................................................................ 53

REFERENCES .......................................................................................................... 54
Table of Contents—Continued

APPENDICES

A. R-CBM Directions and Scoring from the DIBELS Next Manual ...................... 60
B. Human Subjects Institutional Review Board Letter of Approval ...................... 63
C. Teacher Consent................................................................................................. 65
D. Parent Notification .............................................................................................. 69
E. Student Assent Form........................................................................................... 71
F. Example of Student Materials for an R-CBM Probe ........................................ 73
G. Example Scored R-CBM Reading Probe............................................................ 75
H. AIMSweb Passages Randomly Selected for This Study...................................... 77
I. R-CBM Self-Assessment Accuracy Checklist...................................................... 79
J. R-CBM Procedural Integrity Assessment Accuracy Checklist............................. 81
K. Example Teacher Report of Study Data............................................................... 83
LIST OF TABLES

1. Summary of Readability Formulas ................................................................. 12
2. Mean, Deviation, and Mean Euclidean Distance by Passage .............................. 43
3. Estimate of Variance Components.................................................................... 45
LIST OF FIGURES

1. Deviation scores (wc/min) on the 20 passages read from DIBELS Next (top graph) and AIMSweb (bottom graph)........................................................................................................ 44
CHAPTER I

REVIEW OF THE LITERATURE

Introduction to CBM

Curriculum-based measurement (CBM) is a method of assessment that measures student growth in the school curriculum (Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994). The focus of CBM is to assess student progress in core academic subjects, such as reading or mathematics (Fuchs & Deno, 1992). Deno (2003) points out that CBM originated as a way to measure student achievement repeatedly to evaluate instruction and improve effectiveness. To accomplish this, a CBM is made up of sample tasks representative of the curriculum and corresponding to desired year-end performance. Student behavior is evaluated at regular intervals during the school year using equivalent test forms, and the results are graphed over time to establish a slope or trend line for an individual student (Fuchs, 2004; Fuchs, Fuchs, & Compton, 2004).

The skills assessed through CBM are foundational requisite skills that students must acquire to achieve satisfactory academic progress in subjects such as reading, mathematics, spelling, and written expression. The goal when using CBM is to assist teachers in knowing if a student has mastered these critical skills. Hosp, Hosp, and Howell (2007) outline the key features of CBM. Hosp et al. state that the assessments must be “(a) aligned with curriculum, (b) sensitive to instruction, (c) repeatable so that progress monitoring can occur, and (d) criterion referenced so that they could be used to
determine when a student has mastered a task” (pp. 7-8). CBMs are used in schools to identify those students who are behind and need intervention, to progress monitor student growth, and to evaluate instructional effectiveness.

Introduction to R-CBM

Reading curriculum-based measurement (R-CBM) is a widely used method in education for evaluating overall reading achievement at the group and student level. Fuchs (2004) conceptualizes R-CBM as a valid alternative to mastery measurement and other academic forms of monitoring student progress in a curriculum. R-CBM is a brief fluency measure found to be a reliable and valid indicator of overall student performance in reading (Deno, 1985; Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Good & Kaminski, 2002; Hosp & Fuchs, 2005; Rouse & Fantuzzo, 2006; Shinn, 1989, 1998; Wheldall & Madelaine, 1997). Teachers and other educational professionals use R-CBM to track accuracy by calculating how many words correct per minute (wc/min) a student reads. To complete R-CBM, a teacher conducts a 1-minute timed assessment with a student in which the student reads a text passage. The teacher monitors how many words are read correctly and how many errors occur throughout the 1-minute assessment. A word read correctly is defined as a word pronounced correctly within the context of the sentence, while mispronunciations, reversals, omissions, and hesitations of more than 3 seconds are considered errors (Hosp et al., 2007; Shinn, 1989). A more detailed discussion of how to calculate words read correctly is found in Appendix A. Additionally, Shinn and Shinn (2004) have a manual developed describing how to administer and score R-CBM.
Reading curriculum-based measurement is also classified as a general outcomes measure (GOM) indicating that this one assessment is a measure of a long-term capstone skill, rather than a short-term objective. Kaminiski and Cummings (2007) describe GOM as asking, “Is the student learning and making progress toward the long-term goal?” (p. 27). General outcome measures assess critical skills that are representative of important overall outcomes. In reading, the long-term goal in using any outcome measure is to ascertain whether over time the student is able to perform with less effort and more accuracy. Therefore, a reading GOM includes multiple versions of a similar assessment where students demonstrate how fluently they are able to move through a reading passage. Fluency in reading is considered a GOM because to read fluently, or effortlessly, a student must be able to use a myriad of reading skills at the same time. These include decoding words, prediction, vocabulary understanding, syntax, and overall comprehension of the story (Hosp et al., 2007). Deno, Fuchs, Marston, and Shin (2001) point out that, due to the use of a variety of assessment approaches converging in this measure of fluency, R-CBM is able to concurrently provide information regarding how a student is presently functioning as well as whether a student is making gains across time. By using standardized administration and scoring procedures, as well as measuring a broad concept, R-CBM can give an overall indication of how a student is progressing in reading. The assessment is developed in a way that passages of presumed comparable difficulty are repeatedly utilized over a period of time to determine if student progress is being made.
Practical Uses of R-CBM

R-CBM was first developed for the purpose of relatively low-stakes decision making regarding effectiveness of interventions with students, specifically those students in special education programming (Deno, 2003). Increasingly, school professionals working with all categories of students, including general education students, have been called upon to demonstrate that the services provided are effective. The expectation by policy makers, legislators, and those in the community is that decisions made at the systems, grade, and student level will be based on data, and that the efficacy of instruction can be demonstrated. The National Reading First Assessment Committee (Kame’enui, 2002) stated that school-wide reading assessments should be able to screen students for reading difficulty, progress monitor to determine necessary modifications to student programming, help teachers plan what instruction is needed, and give an idea of what types of outcomes are occurring with current instruction at the student and grade level. Deno’s (2003) comprehensive review of the use and practices of CBM assures that R-CBM can be used for all of these purposes, as well as assisting schools in predicting performance on high stakes testing, developing local norms, and increasing communication between staff, parents, and students.

Research has demonstrated the usefulness of R-CBM in describing overall reading achievement (Fuchs & Deno, 1992; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shapiro, 1996), in tracking progress of students across time and tracking effectiveness of interventions (Deno & Fuchs, 1987; Skinner, Belfiore, & Watson, 2002), and even in tracking the effectiveness of school-wide curriculum (Good, Gruba, &
Kaminski, 2002). Studies have shown that when teachers utilize R-CBM to track student progress, students are more likely to make academic gains (Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 1986; Fuchs, Fuchs, Hamlett, & Ferguson, 1992), and those who work with the student are more accurate regarding student achievement levels (Fuchs et al., 1984; Fuchs & Fuchs, 1992; Madelaine & Wheldall, 2005). Research has indicated that R-CBM is extremely sensitive to academic growth and, therefore, is a measure that is useful when monitoring student progress (Fuchs, Fuchs, Hamlett, et al., 1994). Continuous samples of student performance using R-CBM enable teachers to determine whether students are on par with average or high-achieving peers, as well as whether classroom instruction and intervention is moving the student toward achievement of academic goals.

**R-CBM in the Response to Intervention (RtI) Model**

Special educators and school psychologists have long raised measurement issues related to the traditional model of learning disability (LD) assessment for special education eligibility. The IQ-achievement discrepancy model, the traditional form of assessment for LD eligibility, has been hotly contested since its inception (e.g., Algozzine & Ysseldyke, 1987; Learning Disabilities Roundtable, 2002). Researchers such as Algozzine and Ysseldyke identified serious flaws in the IQ-achievement model, which led, in part, to research on alternative assessment options. The discussion of LD assessment has largely turned to a Response to Intervention (RtI) model that utilizes continuous monitoring of student achievement in the general curriculum, and then with
intervention, to determine whether students are making adequate progress or whether they may need special education programming to be able to move forward academically.

Due to the ability R-CBM has to measure both the critical information of current student achievement and progress toward academic goals, R-CBM has been highly discussed in the RtI assessment literature. RtI is a system of assessment and intervention that has a primary objective of preventing academic failure. Early intervention and prevention can change student trajectories and move students to a path where they will be able to be academically successful (Fuchs et al., 2004). The overall goal of an RtI model is to catch students who are struggling to meet basic educational standards and to identify and implement research-based strategies and interventions to directly assess students’ needs in a proactive manner. The purpose of assessment in an RtI model is screening of all students to monitor and identify students early who are struggling, and monitoring the progress of those who are behind to monitor instructional effectiveness.

As RtI has taken shape, R-CBM has played a role in the early identification process as well as monitoring whether students are making gains toward goals. R-CBM is not only a popular way to identify those in need of additional intervention support, but also a way to track progress. Fuchs and Fuchs (1998) outline a RtI dual-discrepancy model that emphasizes the use of R-CBM to define problems and evaluate the student response to implemented interventions. A piece of the dual discrepancy model refers to collecting data that indicate whether students are achieving at a level that is significantly discrepant from their peers. The use of local norms provided through universal screening with R-CBM is routinely used in schools to determine whether students are making adequate progress and meeting critical reading goals. Once individual student goals are
set, if students are not meeting goals, then they may be considered discrepant from expected progress.

This dual-discrepancy model inherently leads to the need for progress monitoring. The RtI model is moving school teams toward making instruction, intervention, and even special education eligibility decisions based on progress monitoring data rather than intelligence and traditional norm-referenced achievement tests. Instead, schools are using screening and progress monitoring data, such as R-CBM, to assist in the eligibility determination process. Within this model, rate (or slope) of student growth becomes the indicator regarding effectiveness of intervention. Due to the increased and high-stakes nature of the role that R-CBM is playing, good quality progress monitoring is essential. Assessments used must hold to high standards of technical adequacy, specifically reliability and validity.

The National Center on Response to Intervention (2010) defines progress monitoring as:

Repeated measurement of academic performance to inform instruction of individual students in general and special education in grades K-8. It is conducted at least monthly to (a) estimate rates of improvement, (b) identify students who are not demonstrating adequate progress, and/or (c) compare the efficacy of different forms of instruction to design more effective, individualized instruction. (para. 2)

The National Center on Response to Intervention has created a list of tools reviewed by a committee of professionals highly knowledgeable in the field of progress monitoring. This list gives suggestions regarding assessments to use in the RtI process. A review of this list by the lead author on August 14, 2010 indicated that of the assessments identified by this group, six out of the eight recommended reading tools utilize R-CBM procedures.
It is clear that R-CBM is recommended as an instrument for determining student achievement, growth, and discrepancies as well as sanctioned to be used as a guide for evaluating progress toward academic goals.

Research in R-CBM

Due to the increased applications of R-CBM and especially due to the interest in using R-CBM as a part of the assessment for eligibility process, the use, technical adequacy, and factors that affect assessment outcomes must be fully analyzed (Fuchs et al., 2004). These assessments must provide solid, dependable information to teachers and other educational professionals. Deno (2003) and Fuchs (2004) outlined three stages of the progress of research on R-CBM: (a) evaluation of the technical features of the resulting static score, (b) determination of the technical features of slope, and (c) assessment of the utility of the measure in the instructional setting. Fuchs indicates that examination of the technical features of progress over time, measured as slope of improvement, is a critical second stage given that one of the key features of R-CBM is its capacity to demonstrate student growth and response to intervention. Relatively little research has examined R-CBM slope, however, perhaps because this research is intricate and time consuming (Fuchs, 2004).

It is clear that R-CBM is recommended as a tool for determining student achievement, growth, and discrepancies as well as to guide the recommendation of students for possible special education eligibility. Repeated assessment of student achievement is a valid and necessary piece of RtI eligibility determination. Fuchs et al. (1984) demonstrated successfully that teachers who used a repeated measurement
technique to monitor instruction and intervention had overall greater success with their students than those teachers who did not use a formal monitoring process. In fact, this study found that students who were being monitored with R-CBM showed significantly larger increases in decoding, fluency, and comprehension skills than their counterparts in classrooms without frequent assessment. Teachers who used this model were better able to describe student achievement and progress; students were more knowledgeable about their own level of achievement. Moreover, all of those involved found the process to be beneficial and meaningful. The indication is that students and school communities are benefiting from the use of frequent measurement through R-CBM. However, debates continue over a variety of aspects of the assessment.

As the use of R-CBM for overall student monitoring of academic growth has increased, so has the interest regarding variables other than student learning that may show variations in assessment results. Dunn and Eckert (2002) investigated whether it is better to progress monitor with material at the student’s instructional level, or at a more challenging level. Results showed that while the two methods can have similar results regarding overall gains in reading, there was considerable variance in the data that was unexplained by student growth, achievement, or characteristics. Studies have set out to investigate the variables responsible for such variance and several statistically significant influencers have been identified. Christ (2006) points out the importance of testing conditions and its effect on resulting student scores. Christ found that poorly controlled assessment conditions generated “four times more error than optimally controlled conditions” (p. 131). Other factors that can affect the variability of R-CBM include who administers the assessment and where it is administered (Christ, 2006; Derr & Shapiro,
1989; Derr-Minneci & Shapiro, 1992), the type of probes that are used (Fuchs & Deno, 1992; Hintze, Daly, & Shapiro, 1998; Shinn, Gleason, & Tindal, 1989), and the specific directions given to students (Colon & Kranzler, 2006).

**R-CBM Readability**

The readability of the passages that are used with students is an important determinant of the variance of R-CBM data. Deno (2003) indicates that one of the most critical aspects of quality R-CBM is that the assessment is conducted with equivalent, but novel, passages. To be able to use R-CBM and progress monitoring as intended, passages must have consistent difficulty level and present as parallel forms across the grade level.

In a position paper written by Kaminski, Cummings, Powell-Smith, and Good (2008), the authors report that R-CBM is:

> Sensitive to small, but important changes in student performance. . . . Differences in scores are attributable to student growth, not differences in the materials or assessment procedures so educators can compare assessment results over time. (p. 1188)

The general supposition has been that due to the use of readability formulas to equate passages, variation in student achievement on R-CBM not explained by student achievement differences was due to error associated with constructs (i.e., variables) such as motivation level, variation in background knowledge, and testing conditions. Differences in passage difficulty were not considered as a source of error due to the use of readability estimates to develop equivalent passages.

Developers of R-CBM packages have worked to use a variety of readability formulas in an effort to decrease the factor of difficulty level of the passage affecting
student scores. The central reason for developing homogenous passages revolves around the assumption that R-CBM is an enhanced assessment system because of its ability to monitor with equivalent materials. If the assessments are not equivalent, the student’s progress with reading may be miscalculated. If a teacher sees an increase in words correct per minute, it is assumed that learning has occurred. However, if the most recently read passage was simply easier than the previous passage, then this assumption may not be correct. If reading passages are not equivalent, changes in performance may reflect passage difficulty level of the individual passage rather than changes in actual student achievement. While the developers of R-CBM passage sets have used readability formulas, further investigation of R-CBM has indicated that the traditional readability formulas may not have sufficed for R-CBM passage equating.

Readability is a complex construct, and while those in the field of developing formulas to level text have worked hard to make this into a science, much that may affect readability is not considered in the formulas (Bailin & Grafstein, 2001). Investigations of how various readability formulas relate to the measurement of oral reading fluency have begun, but the results are not conclusive. It is critical to understand from the outset that the most popular readability formulas have been developed to link difficulty of text to the ability to comprehend written text (Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005; Bailin & Grafstein, 2001). It is increasingly apparent that the traditional methods of readability determination may not wholly translate to oral reading. In addition, each popularly utilized readability formula includes different components to determine the difficulty level of the text. There is not a certain standard or best practice way to
determine the difficulty level of text. This is illustrated in Table 1, which summarizes commonly used readability formulas.

Table 1

*Summary of Readability Formulas*

<table>
<thead>
<tr>
<th>Readability Formula</th>
<th>Developer</th>
<th>Date Developed</th>
<th>Appropriate For</th>
<th>Components Taken Into Account</th>
</tr>
</thead>
</table>
| Dale-Chall\textsuperscript{a} | Edgar Dale and Jeanne Chall | 1948 | Above 4\textsuperscript{th} grade | – Average sentence length  
– Number of unique unfamiliar words |
| Flesch-Kincaid\textsuperscript{a} | Rudolph Flesch | 1948 | All levels | – Word length  
– Sentence length |
| FOG\textsuperscript{a} | Robert Gunning | 1968 | Adult text such as newspapers | – Average sentence length  
– Number of complex words (3 or more syllables) |
| Forecast\textsuperscript{a} | John Caylor, Thomas Sticht, and J. Patrick Ford Commissioned by U.S. Army | 1973 | Not for running narrative. Better used with multiple choice tests. | – Number of monosyllables |
| Fry\textsuperscript{a} | Edward Fry | 1968 | All levels | – Number of sentences  
– Number of syllables |
| Powers-Sumner-Kearl\textsuperscript{a} | R. D. Powers, W. A. Sumner, B. E. Kearl | 1958 | Ages 7-10 | – Average sentence length  
– Percentage of monosyllables |
| SMOG\textsuperscript{a} | McLaughlin | 1969 | 4\textsuperscript{th} grade and above | – Number of words with 3+ syllables |
| Spache\textsuperscript{a} | George Spache | 1953 | 3\textsuperscript{rd} grade and below | – Average sentence length  
– Number of unique unfamiliar words |
| DMG Passage Difficulty Index\textsuperscript{b} | Cummings, Wallin, Good, Kaminiski | 2007 | Not reported. Used with Grades 1-6 DIBELS Next Passage Set | – Characters per word  
– Proportion of words with 7 or more characters  
– Syllables per word  
– Proportion of words with 3 or more syllables  
– Proportion of rare word  
– Words per sentence |

\textsuperscript{a} “Readability Formulas” (2010)  
\textsuperscript{b} Powell-Smith, Good, and Atkins (2010)
At the beginning of the use of R-CBM to monitor students, the difficulty level of the text was not carefully considered. Shinn (1989) suggested that the passages should be equivalent, and he discussed the use of readability formulas. Typically though, during the early years of reading curriculum-based measurement, R-CBM passages were developed from basal reading series and considered equivalent because the reading text publishers utilized readability formulas. As more research on CBM was done, investigators began to highlight the fact that difficulty level of passages chosen for progress monitoring can have significant effect on resulting R-CBM scores (Hintze et al., 1998; Hintze, Owen, Shapiro, & Daly, 2000). While using assessment passages developed straight from the reading curriculum did seem to have face validity for teachers (Fuchs & Deno, 1994), passages drawn from the classroom curriculum have much variability in readability (Fuchs, Fuchs, & Deno, 1982). Today the common suggestion is the use of prepackaged passage sets whose readability has already been further investigated by the publisher of the passage set (Hosp et al., 2007).

Passage sets with more equivalent difficulty levels are an improvement in R-CBM. Research by Hintze and Christ (2004) demonstrated that controlling the level of difficulty in the reading text, as developed passage sets do, can lead to assessors minimizing the effects of passage or text difficulty as a potential source of measurement error in relation to those passages created from a basal series. Further investigation indicates that while passage sets equated through readability formulas may be better than passages sets not developed in this way, the former passages may not be as comparable as originally thought. Hintze and Christ reported that the standard error of the slope (SE\(b\)) of the passages reduced from 1.27 for an uncontrolled probe set to 1.07 when controlling for
passage difficulty through use of created passage sets utilizing a readability formula. While this is a meaningful decrease, much error related to passage difficulty continues to be a part of the R-CBM assessment process.

Betts, Pickart, and Heistad (2009) reviewed the technical manual of a popular R-CBM passage set, AIMSweb (Howe & Shinn, 2002). Betts et al. found that the often-used Spache readability formula did not correspond well with the resulting student scores on the passages. The assumption was that if the readability formula showed a text to be slightly more difficult, students should score lower on the passage. This hypothesis did not hold true when Betts et al. investigated student data reported in the AIMSweb technical manual. Betts et al. continued this research through field testing of passages and found that readability formulas were able to establish differences between passages of differing grade levels, but did not seem to be able to distinguish well the difficulty level within a grade level. Additional studies investigating the use of readability formulas by Francis, Santi, Barr, Fletcher, Varisco, and Foorman (2008) and Ardoin, Williams, Christ, Klubnik, and Wellborn (2010) concur that the use of readability formulas are not equating passages to a level that is adequate for the types of decisions being made with R-CBM. These studies indicate that while readability may help begin the process of passage selection, popularly used readability formulas do not seem to provide form equivalence within a grade level. Ardoin, Williams, et al. did find that student scores on passages were significantly related to each other. In other words, passages which were relatively easy for a student were typically easier for the other students who read the same passages. Ardoin, Williams, et al. point out that this indicates that passage difficulty is not individual to each student and therefore is not a random construct.
Ardoin, Suldo, et al. (2005) took on the task of investigating more closely how R-CBM passages hold up to a variety of well known readability formulas as well as common variables related to readability. Their research indicated that five readability formulas were the most commonly used in R-CBM development: Dale-Chall, Flesch-Kinkaid, Fog, Fry, and Spache. They went on to dig deeper into the relationship between eight reliability formulas as predictors of student’s resulting score, the five most widely used along with Forecast, Powers-Sumner-Kearl, and SMO. Results indicated that there was not a significant relationship between student results and determined difficulty level based on the most commonly used readability formulas. The Forecast and the FOG readability formulas were the top two predictors, with only moderate correlation. Results showed that Spache and Dale-Chall were the worst two predictors. Ardoin, Suldo, et al. concluded that there is not research support behind using these readability formulas in determining difficulty level of passages related to oral reading. Components of a reading passage that were found to correlate with student results were found to be the number of syllables per 100 words, and number of words on the Dale-Chall 3,000 word list.

Ardoin, Suldo, et al. (2005) strongly suggest the need for a better formula in developing reading probes for R-CBM. The authors also propose the practice of giving the probes to a large number of students and taking out those probes that have a high level of variance. Best practice in developing packaged progress monitoring sets would include giving the created passages to a group of test students and taking out the assessments that are found to be much easier or much harder for students to read orally, leaving only those passages that were found to be of equivalent difficulty level in the test group condition. Ardoin, Suldo, et al. suggest that field testing and deleting passages with high variance
“may be the best and only valid measure of the difficulty of a passage when evaluating fluency” (p. 18).

The characteristics of readability were also investigated by Compton, Appleton, and Hosp (2004), who found that levels of difficulty yielded by two popular readability formulas, Felsch-Kincaid and Spache, were not correlated with each other, with a resulting correlation coefficient of .28, indicating that the two readability formulas are measuring different constructs. This study also found that passage accuracy and fluency were significantly associated with the percentage of high frequency words in the passage, and fluency was well correlated with the ability to decode words within the passage. Compton et al. concluded that passages with lower word recognition demands, and greater numbers of high frequency and decodable words were generally read with greater accuracy and fluency in developing readers. Although increased high frequency words assisted both low and average readers in more fluent reading, percentage of decodable words in the passage did not increase accuracy and fluency of low readers.

Poncy, Skinner, and Axtell (2005) examined further how much error is associated with differences in passage difficulty. When investigating a passage set developed by Good and Kaminski (2002), Dynamic Indicators of Basic Early Literacy Skills (DIBELS) version 6.0, results indicated that 81% of the variance was due to actual differences in students’ reading skill, 10% was from passage variability, and 9% was attributed to unexplained error. Poncy et al. then tested 20 passages with each of the 37 student participants and determined which passages were comparable in level of difficulty for each student. When this set of better equated passages for each student was analyzed, error based on passage variability was reduced to just 1% and the amount of variance
explained by the student increased to 89%. Research projects have since taken on this type of process to create progress monitoring passage sets that have low variance for each individual student (e.g., Daly, Persampieri, McCurdy, & Gortmaker, 2005). However, the time necessary to develop a homogenous passage set for each student assessed for reading problems in schools makes this process seem neither practical nor feasible.

Rather than developing a separate equivalent passage set for each individual student, Christ and Ardoin (2009) set out to find a better method of equating developed R-CBM passage sets. Four methods of developing passage sets were used: (a) random selection of passages from a reading curriculum using the procedures set forth by Shapiro (1996), (b) readability formula utilized to find those within a certain readability range, (c) using the performance means after field testing with students to determine more equitable passages, and (d) Euclidean distance procedure used to group similar passages and find outliers to discard from the passage set. Results indicated that using the Euclidean distance method greatly decreased error associated with variance in passage difficulty. The Mean Euclidean Distance (MED) is the square root of the sum of squared differences between repeated measurements (Christ & Ardoin, 2009). This procedure groups observations together to identify outliers. When analyzing MED, passages with lower MED magnitudes are considered more stable passages since those with lower levels produced fewer outliers.

In a subsequent study when a passage set developed using Euclidean distance was compared to two popular passage sets, DIBELS version 6.0 and AIMSweb, results indicated that the passage set developed through the Euclidean distance method showed decreased measurement error and that the AIMSweb passages had less measurement error.
than the DIBELS version 6.0 passages (Ardoin & Christ, 2009). This study may shed light on an improved method of developing a passage set that has more homogeneous passages.

R-CBM Slope

When determining whether students are making appropriate gains in reading, overall trend across time is investigated. If students are making adequate upward gains in achievement toward their set goal, then instruction and intervention are considered successful. When this trend is investigated, using graphed data indicating student progress, such as R-CBM measures, the slope of improvement over time is the primary index of success (Hintze & Shapiro, 1997). Fuchs (2004) explains that because the goal of R-CBM is that each passage used is comparable in difficulty, slope can be used to determine the progress of a student. Fuchs goes on to state that “slope can also be used to gauge a student’s responsiveness to the instructional program and as a signal to revise the student’s program when inadequate responsiveness is revealed” (p. 189). To have confidence that slope reflects learning, however, the passages must be parallel, that is, of consistent difficulty level. There must be low variance between each assessment probe to be able to make individual student decisions regarding growth trajectories.

Studies have accomplished the task of determining what type of slope growth should be expected using R-CBM; however, these studies are investigating group level data rather than individual differences across time. Deno et al. (2001) suggested that a one-word gain per week in words correct per minute (wc/min) is an average expected gain and that even higher rates of improvement should be expected with strong
instructional practices. Deno et al. did find that investigating each grade level expected growth independently, as well as having different growth estimates for those qualified for special education services, may be appropriate. Hintze, Shapiro, and Lutz (1994) also found that the mean weekly slope of third graders was around 1 wc/min increase per week. Unfortunately, a high amount of error variance was found with this sample, leaving practitioners to wonder about the source of this variance and concluding that it would be impossible to make meaningful individual student decisions. Silberglitt and Hintze (2007) found that there is variance in how much growth can be expected across time based on initial R-CBM performance. Most notably, those students who were considerably behind in reading did not make the same growth as average performing students.

There is common consensus that when looking at long-term progress, such as a period of 1 to 2 years of assessment data, error related to the results and analysis are low (Deno et al., 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). While it is positive that long-term investigation of progress monitoring data leads to a strong and accurate depiction of student performance, in school settings a quicker view of progress is much needed, and much expected, from assessments. Good and Shinn (1990) and Shinn (2002) suggested that a minimum of 10 data points is needed for high-quality decision making regarding progress; however, when researchers have looked at assessment periods such as 10 or 12 weeks of progress monitoring, considerable measurement error relative to slope was still found (Christ, 2003; Hintze et al., 1998; Hintze & Shapiro, 1997). Christ (2003) reports that while there are reports of how much student growth can be expected at each grade level when monitoring with R-CBM, results are not so straightforward when the
confidence interval is also reported. When describing a student’s wc/min gain per week, a confidence interval may indicate that almost none, to quite a bit of growth has occurred. Christ argued that when confidence intervals are investigated, results make it impossible for a practitioner to know if there was no progress at all, or highly significant progress.

The standard error of the slope ($SE_b$) provides information regarding the error associated with resulting regression slopes. Studies by Christ (2006) and Francis et al. (2008) have pointed out that error associated with slope of individual student progress monitoring often is large enough to result in reservations regarding the interpretation of rate of growth. They noted, however, that $SE_b$ is smaller the longer the duration of monitoring. As Christ (2006) pointed out, the median $SE_b$ fell from a median of approximately 9.19 wc/min for 2 weeks of monitoring, to 2.21 for 5 weeks of monitoring, and finally to 0.42 when 15 weeks of progress monitoring was observed. Christ and Choolong-Chaffin (2007) suggested reporting standard error of measurement (SEM) when R-CBM results are conveyed for better understanding of true fluency score, especially when multiple administrations of the assessment are not available. Christ (2006) and Christ and Silberglitt (2007) suggested reporting $SE_b$ and possibly a confidence interval when communicating R-CBM progress monitoring results.

Francis et al. (2008) found that the effect of various difficulty levels of the passages altered the shape of growth trajectories and affected estimates of slope. The students’ scores indicated that fluency rates varied considerably within passages, as well as between passages. Through investigation of the raw data, it was evident which passage was easiest and which was more difficult for the students. While they found that R-CBM does well at assessing oral reading fluency, it was not evident that the passages within a
progress monitoring set could be considered equivalent passages. They concluded that there are large disparities in student scores based on the use of different passages that are apparently not well equated. Francis et al. pointed out that:

> Teachers run the risk of misinterpreting gains and losses in [wc/min] as gains and losses in student fluency when, in fact, they are little more than expected differences resulting from the use of easier or more difficult passages from one week to the next. (p. 334)

Francis et al. then utilized a method for equating passages and reanalyzed the data. Once passages were equated, the error associated with passage difficulty was eliminated and better decisions regarding growth could be made.

Ardoin and Christ (2008) focused on the slope of three times a year monitoring of student progress. The study investigated the effect of using alternate passages at each assessment period that were equated using readability estimates, using one same passage at each assessment period, and using the same three-probe passage set at each assessment period with second grade students. Results showed that, for second grade, students showed more academic growth in the first semester of the year than in the second semester. A high number of students assessed did not meet the 1 wc/min expectation set out by Deno et al. (2001). In regards to the issue of slope estimates, the study showed that the most reliable slope estimate was found when using the same passage set at each measurement period and the median score. This procedure decreased error variance, possibly error from variance in difficulty level of the assessment passages. Ardoin and Christ (2008) also addressed the issue of practice effects and demonstrated that with this sample practice effects were not an issue. A caveat to this is that the sample was used
with second-grade students. An argument could be made that memory capability, and therefore practice effects, may be different for students at a higher grade level.

In a recent article by Ardoin and Christ (2009), the authors compared the growth estimates of three progress monitoring passage sets: (a) Formative Assessment Instrumentation and Procedures for Reading (FAIP-R), a passage set created by Christ and Ardoin (2009); (b) AIMSweb and (c) DIBELS version 6.0. AIMSweb and DIBELS version 6.0 were included due to their popular use in the educational community. The results showed that the FAIR-P passages had significantly less measurement error than did the other passage sets and that AIMSweb had significantly less measurement error when compared to DIBELS version 6.0. The resulting differences were thought by Ardoin and Christ (2009) to be due to varying strategies used when determining readability estimates of passages to include in each set. All passages utilized readability formulas for initial development. AIMSweb included a procedure of using a sample group to pilot the passages and took out passages that showed high level of variance within this sample (Howe & Shinn, 2002). FAIR-P took this process a step further and used the Euclidean distance method to eliminate passages that showed high variance with a sample population (Ardoin & Christ, 2009). The authors contended that Euclidean distance is a superior method of determining variance and that utilizing this procedure led to better equated passages. This study is the first to examine the differences between AIMSweb and DIBELS version 6.0 and suggests that the quality of the passages vary considerably and that the “AIMSweb passage set resulted in significantly smaller $SEb$ and SEE” (Ardoin & Christ, 2009, p. 278).
Overall, it is evident that there is error associated with student score results when using R-CBM based on the variability of the passage difficulty. The investigation of the slope across time using R-CBM assessments may also be affected by the variations in the difficulty level of the passages. To address these issues, one suggestion is to use the same probe set for each assessment. This may be appropriate for three times a year benchmarking, but has not been demonstrated as appropriate for regular progress monitoring which may occur on a biweekly or even weekly basis. Poncy et al. (2005) demonstrated that using a system of initial assessment with a student to determine which passages are of equal difficulty for that student and using these passages to progress monitor is effective in decreasing error related to variability. While this does seem to lead to a passage set that is more homogenous for that student, it is not a feasible practice in an applied setting. Still others who have investigated readability procedures have suggested using different procedures to determine R-CBM passage difficulty equivalence, such as using the Forecast readability formula, using percentage of monosyllables, using the Dale-Chall 3,000 word list, and field testing and removing passages with large error variance to increase the equivalence of created passage sets.

The resulting analysis of the body of research regarding the variability in the readability of R-CBM passages and the resulting seemingly lack of ability to make a strong statement of level of progress and growth of student achievement across time makes it difficult to determine the utility of current R-CBM data at the individual student level. It is clear that it is necessary to equate the readability of R-CBM passages and that doing so will increase the accuracy of progress monitoring data and allow for improved and more reliable predictions of student growth. Those in the field have been calling for
new and better ways to determine equivalent passages. Continued research is needed as to whether it is appropriate to make statements regarding whether students are making adequate academic progress based on R-CBM progress monitoring data.

DIBELS Version 6.0 and DIBELS Next Passage Development

The DIBELS website (https://dibels.uoregon.edu/) reports that their data management system has been used in over 15,000 schools. Countless other schools and classrooms have utilized the passage sets provided by DIBELS through their free download offering. At the time that DIBELS version 6.0 was developed, readability was considered and was a part of the overall development of the passages used. Passages were written to target end-of-year difficulty for each grade level except for first grade, which targeted beginning of the second grade year difficulty (Good & Kaminiski, 2002). Use of readability formulas was the exclusive method used to determine whether passages were of equivalent difficulty. The Spache readability formula was used for each grade level; however, an investigation of the technical report related to passage development indicates that when other readability formulas were applied to the passages, they showed high variability in difficulty range (see Table 4 in Good & Kaminski, 2002, p. 7).

DIBELS Next is a new passage set developed by the Dynamic Measurement Group (DMG). Powell-Smith, Good, and Atkins (2010) report that the process used to equate passages for this set is improved and takes into account much of the research base that has explored the subject of R-CBM readability in the last 20 years. To begin with, passages were developed and then analyzed with the DMG Passage Difficulty Index rather than relying on traditional readability formulas which have been found to be
inadequate for equating R-CBM passages. This DMG Passage Difficulty Index, developed by Cummings, Wallin, Good, and Kaminski (as cited in Powell-Smith et al., 2010), examines three aspects of passage difficulty: word difficulty, semantic difficulty, and syntactic difficulty. The formula involves looking at each feature independently and then as an overall composite so as to keep each resulting difficulty reported within a specified range (Powell-Smith et al., 2010). After 40 passages were selected for each grade level based on the readability formula, the passages were field tested with 22-25 students at each grade level. This procedure took into account research by Ardoin, Suldo, et al. (2005) that indicated that field testing passages is essential in the construction of a homogenous passage set. Once the group of students read the passages, they were analyzed looking for outliers. Outliers were defined as passages that were found to be significantly more difficult or easier for students than expected based on the readability results. After the data were analyzed, the best 32 passages per grade, with the exception of 29 for first grade, were selected for use in the DIBELS Next passage set.

Once the best developed passages were determined, the developers of DIBELS Next organized passages into triads. While it is reported that the most homogenous passages were chosen for this passage set, the triads are organized to have each of three difficulty levels. Powell-Smith et al. (2010) described the passages in the triads as, “slightly easier . . . of middle difficulty . . . and . . . slightly harder” (p. 15). The authors of DIBLES Next were careful to point out that the differences between passages appear to be slight but are enough to organize each triad with slightly different levels of each type of passage. Also, one progress monitoring dyad was developed for each grade level so as to result in 20 progress monitoring passages. Another piece of the development was to be
sure that each triad had a certain balance of genre. Powell-Smith et al. reported that in first through third grade each triad consists of two narrative passages and one expository passage. In fourth through sixth grade, each triad consists of one narrative and two expository passages. When results were analyzed based on the triads, it was found that there was “most typically a mean difference of about 2 words read correct from triad to triad” (Powell-Smith et al., 2010, p. 20). The developers contended that this most recently developed passage set more thoroughly controls for passage difficulty and that the arrangement of passages in triads better controls for passage difficulty.

It is important to note that within the data analysis conducted to determine the best passages for DIBELS Next and to better equate passages, there were students whose resulting data indicated such high level of variance that their information was not included in the final analysis. Powell-Smith et al. (2010) specifically discussed one student in the third grade whose results were so variable that they suggested caution in making statements regarding progress based on R-CBM progress monitoring without further investigation.

AIMSweb Passage Development

In the 2008-2009 school year, the AIMSweb online data system reports that over 3.5 million R-CBM AIMSweb assessments were given (T. J. Ryan, personal communication, July 16, 2010). AIMSweb passages were developed in a similar fashion to DIBELS version 6.0 in that there was a heavy reliance on readability formulas. Howe and Shinn (2002) described the passage development and readability equating process in a paper available through AIMSweb. Initially, passages were developed by trained
authors and assessed with the Fry readability formula to determine if the passage was written at the appropriate difficulty level. After development, passages were also checked with the Lexile-grade standards, and passages not within a pre-set acceptable range were removed. Initially 33 passages were developed for first grade and 50 were developed for second through eighth grades. Passages were then field tested with 24 students per grade, except for sixth grade, which had 18 students, with a representative sample from the average, above average, and below average reading performance level. Passages that had large variance and were therefore deemed too easy or too difficult to be included in the passage set were eliminated. Eighth grade did show too much variability with the initially developed passages, so at this grade level new passages were written and then field tested again. Resulting passages included 23 for first grade and 33 for second through eighth grades. Howe and Shinn reported alternate form reliability for the majority of the passages as being above .85 and all passages are above .70. SEM values are reported between 6.3 and 13.3 wc/min.

Howe and Shinn (2002) also report the Lexile for each passage. It is interesting to note that there is some overlap in Lexile results between grade levels. For example, the seventh grade benchmark passages have one passage with a Lexile higher than two of the eighth grade passages, and two passages with Lexiles lower than one of the sixth grade reading passages. In sum, this report related to the equating of the AIMSweb passages reports that the technical adequacy of the development process and field testing trial indicates that the assessment is able to be used to make instructional decisions regarding student performance and progress.
Research Questions

Currently there is no agreed-upon method for determining the difficulty level of an R-CBM passage to equate reading passages. For teachers to make correct judgments about student academic progress, it is essential that R-CBM passages are of equal difficulty level. The primary objective in this study is to evaluate the homogeneity of oral reading fluency progress monitoring passages of two popular passage sets that are used frequently in an RtI model of service delivery. The purpose of this research is to examine the stability of each R-CBM progress monitoring passage set. This study investigates the use of DIBELS Next and AIMSweb oral reading fluency measures for the purpose of monitoring student progress in reading. It addresses the following research questions:

1. What is the variance within each passage set, DIBELS Next and AIMSweb?
2. How do the variances from DIBELS Next and AIMSweb compare?
3. How does organizing progress monitoring passages into triads affect variance?
CHAPTER II

METHOD

There is a need to continue to investigate the accuracy and utility of R-CBM progress monitoring. Teachers, schools, and parents are using this technique to determine whether students are benefiting from current instruction, whether instruction needs to be modified, and even how students are progressing through the RtI process. The primary objective of this study is to evaluate the homogeneity of oral reading fluency progress monitoring passages of two popular developed passage sets: DIBELS Next (Good & Kaminski, 2010) and AIMSweb (Howe & Shinn, 2002). The purpose of this research is to examine the stability of the DIBELS Next and AIMSweb R-CBM progress monitoring passage set. This study investigates the use of R-CBM fluency measures for the purpose of monitoring student progress in reading and addresses the following research questions:

1. What is the variance within each passage set, DIBELS Next and AIMSweb?
2. How do the variances from DIBELS Next and AIMSweb compare?
3. How does organizing progress monitoring passages into triads affect variance?

Permission to conduct the study was obtained from the Human Subjects Institutional Review Board of Western Michigan University (Appendix B).
Procedure

Testing sessions occurred daily for eight sessions. At each session the students read five 1-minute timed reading passages resulting in 5 minutes of reading time per session. Allowing for transition time between passages, the testing sessions lasted approximately 8 minutes per session for each student. Sessions occurred in a quiet area close to the students’ classrooms. Those students who were absent for one or two testing sessions were assessed through make-up sessions. If a student missed more than two testing sessions, he or she was removed as a participant.

Participants and Setting

This study was conducted with third grade students in two Parchment School District elementary schools located in Parchment, Michigan. All students in the participating third grade classrooms were initially assessed by school employees, e.g., classroom teachers, in January 2011 through their typical universal screening process using the school’s standard assessment, DIBELS Next. DIBELS Next screening results organizes third grade students into three achievement categories—benchmark, strategic, and intensive, as follows:

1. Benchmark: Reading 86 or more words correct per minute. Students at this level are considered on track with reading and are predicted to do well with reading long-term.

2. Strategic: Reading between 68 and 85 words correct per minute. Students at this level are considered behind in reading fluency and are considered at some
risk for later reading failure without intervention and close monitoring of grade level reading progress.

3. Intensive: Reading 67 words correct per minute or below. Students at this level are considered significantly behind in reading fluency and are considered at risk for later reading failure without significant intervention and close monitoring of reading progress.

Those students who were found to be in the benchmark and strategic reading ranges based on this January school-wide screening assessment were considered for participation in the study. Students who are reading at the benchmark and strategic ranges typically have grade level or above goals in reading. It is appropriate current educational practice to monitor their reading with grade level material. Those students who are found to be reading at the intensive level are considered to be significantly at risk for reading failure and by and large need to be monitored more closely at their instructional reading levels rather than with grade level material. Therefore, including students at this low reading level is not appropriate because it is not standard practice to repeatedly assess with grade level materials that would be potentially too difficult and would frustrate them.

Fifty-four students were initially found to be within the strategic and benchmark reading ranges in the two schools. A further review of those students whose January screening results showed them to be at the benchmark and strategic reading levels was conducted by classroom teachers to eliminate any students with the following characteristics: English as a Second Language (ESL) status, high rates of absenteeism, significant behavior issues that prevent participation in regular classroom activities, and
sensory impairments such as hearing impairment or low visual or blindness. Students with disabilities such as emotional impairment, learning disability, or cognitive impairment were included in the study if they met the other criteria. No students were reported to fit the exclusionary factors, and therefore all 54 students began as participants. Attrition did occur due to absences throughout the assessment period with 9 students being removed from the study due to having three or more absences throughout the testing window. In the end, 45 students, 46.7% male \( (n = 21) \), and 53.3% female \( (n = 24) \) were included in the study. The final group of students was 75.6% White \( (n=34) \), 20% Black \( (n=9) \), and 4.4% Hispanic \( (n=2) \).

In an effort to make students comfortable during the data collection, a number of standard R-CBM procedures were put in place. First, the assessments took place in an area close to their classrooms. This helped to ensure that the child continued to be in close proximity to a familiar adult. Second, children were provided with simple, nonthreatening directions. For example, “I would like you to read a story to me. Please do your best reading. If you do not know a word, I will read the word for you. Keep reading until I say stop. Ready. Begin.” Third, those collecting data were trained that, if at any time during assessments a child appeared uncomfortable or distressed, testing would be discontinued. During the assessment, this never occurred. Students were overall cooperative and seemed to enjoy the individual attention and praise. Fourth, children who were noncompliant or off-task were verbally redirected to tasks, e.g., “Look at this story and do your best reading.” At times students did need to be redirected typically due to the student wanting to talk with the assessor about a story, about reading/school, or about
other topics relevant to the student. No students ever refused participation during the assessment.

Consent and Assent Procedures

The principals of the schools and teachers of the classrooms agreed to have this study conducted with their students. Principals and teachers served as advocates and provided consent for the research to occur in their building and in their classrooms. Those teachers who agreed to participation were presented with oral and written information regarding the project. The project, and its evaluation component, was presented and discussed with principals and teachers collectively at a planning meeting. In addition, a consent form (Appendix C) was presented to the teachers involved with contact information for questions or concerns. All teachers signed the form.

This study involved assessment practices commonly used within the school district and posed little to no risks to all involved. Therefore, parent notification was given rather than collecting parent permission. This parent notification form (Appendix D) was sent out 2 weeks prior to initiation of the study and no one involved in the study reported a parent contacting them regarding the assessments.

Student participants are in the third grade and are of an age where they are able to provide assent. Each student provided assent prior to beginning data collection. Assent was collected through both verbal and written means (Appendix E) prior to the initial progress monitoring session with students. Through the assent process, the students agreed to participate in the study through reading passages as well as agreeing to have their wc/min data results from the R-CBM testing used as a part of this study. On the first
day of testing each student was read an assent form which tells about the study and their participation. When they agreed to participate, they stated so verbally and also wrote their name on the provided form. All students who were eligible in the study agreed to participate.

Materials

*R-CBM Probes*

Twenty passages from each of two probe sets were utilized in this study. These included (a) DIBELS Next third-grade Oral Reading Fluency progress monitoring passages (Good & Kaminiski, 2010), and (b) AIMSweb Reading Curriculum Based Measurement progress monitoring passages (Howe & Shinn, 2002). See Appendix F for an example of student reading material. See Appendix G for an example of a scored R-CBM.

The DIBELS Next passage set was taken from the most current version of the DIBELS materials. The progress monitoring passages for DIBELS Next were developed in groups of six triads and one dyad, where the developers strategically ordered the reading passages in groups so as to result in improved decision making. The procedures and practices outlined by DIBELS Next were honored and the triads and dyad were kept together. The order in which the triads and dyad were read by each student was randomly selected.

The second passage set used was taken from the most current version of AIMSweb R-CBM materials. AIMSweb provides 30 progress monitoring passages. The
passages used in this study were randomly sampled from all of the possible passages. Those selected were read by each student, but the order in which the passages were presented to each student was randomly selected. See Appendix H for a list of which of the 30 third grade passages were randomly selected for this study. The passage sets were counterbalanced so that at each assessment period the passage, DIBELS Next or AIMSweb, read first by the student would vary. All random selection and counterbalancing were determined through the use of a random number generator found at http://randomizer.org.

At each session the data collector said these specific, standard assessment directions prior to the student reading the first passage: “I would like you to read a story to me. Please do your best reading. If you do not know a word, I will read the word for you. Keep reading until I say ‘stop.’ Ready. Begin.” The timing of the passages began when the student said the first word of the passage. Prior to reading the subsequent passages, the data collector said, “Please do your best reading. Ready. Begin.” These directions are standard to R-CBM. See Appendix A for R-CBM Administration and Scoring Directions.

Training Procedures

Assessment was administered by the lead investigator as well as research assistants experienced working with children. Research assistants were college students in the areas of education who have had experience and training working in schools as well as with assessments. Assessors were teachers, former teachers, or college students training to work in a K-12 school setting.
Assessors were trained in R-CBM by the lead investigator. The lead investigator is a DIBELS Mentor Trainer and completed training by Dynamic Measurement Group in the administration and scoring of R-CBM in June of 2010. The lead investigator has been using DIBELS materials in her practice with students since 2001. The lead investigator has led trainings and workshops for teachers in DIBELS version 6.0 and DIBELS Next administration, scoring, and data analysis since January of 2003. She has also been through training provided by AIMSweb regarding administration and scoring of the R-CBM in the summer of 2007 and has been using AIMSweb materials in her practice with students since fall of 2005. Following direct instruction and training of the research assistants, practice administrations of the assessment occurred a minimum of 10 times. Before considering a research assistant to be fully trained and prepared to work with Parchment third grade students, all assistants were able to administer assessment with 100% procedural reliability as evaluated by the lead investigator using the checklist included in Appendix H. Assistants were also able to obtain 100% interrater reliability using point-by-point agreement on at least three consecutive administrations of DIBELS Next and AIMSweb. Interscorer agreement was determined on a word-by-word basis. When both administrators scored a word as a word read correctly, or both scored a word as an error, this is an agreement. When there is a difference in how a word is scored, this is a disagreement. Total percentage of agreements was calculated for each probe by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (House, House, & Campbell, 1981).
Interrater Reliability

Interscorer agreement data were collected on 15% of all possible progress monitoring probes administered during the study. Interscorer agreement was determined on a word-by-word basis. When both administrators score a word as a word read correctly, or both score a word as an error, this is an agreement. When there is a difference in how a word is scored, this is a disagreement. Total percentage of agreements were calculated for each probe by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (House et al., 1981).

Procedural Integrity

Procedural integrity was checked on 100% of the progress monitoring assessments using a self-monitoring checklist. The data collector self-assessed whether they were following procedures on a step-by-step basis. Procedures include whether material was presented in the correct order, directions were followed accurately, time was kept correctly, and overall fidelity of administration was adhered to.

Procedure integrity checks were conducted for fidelity on 15% of the progress monitoring assessments. The integrity checks were conducted by either the lead investigator or a research assistant who is also a DIBELS Mentor. A progress monitoring session conducted by the other data collector was observed. The observer assessed whether the data collector followed procedures on a step-by-step basis using a procedure integrity checklist (Appendix I). Procedures included whether material is presented in the
correct order, directions are followed accurately, time is kept correctly, and overall fidelity of administration is adhered to. Point-by-point agreement was calculated.

Dependent Measures

*Oral Reading Fluency*

The number of words read correctly per minute (wc/min) for each reading passage served as the data collected. Words correct are those pronounced correctly and in the correct order, given the reading passage. Repetitions were ignored and self-corrections were counted as correct. Words not counted as read correctly were mispronunciations, substitutions, and omissions. If a student struggled to pronounce a word or hesitated for 3 seconds, he or she was supplied the word by the examiner and prompted to move on with reading. Such words missed were counted as errors. See Appendix A for further scoring directions. See Appendix G for an example of a scored R-CBM probe.

Risks and Costs

There are no more than minimal risks for student participation in this study, and these risks do not exceed standard risks involved when school officials conduct their R-CBM assessments. For the DIBELS Next and AIMSweb progress monitoring administration, participants left their classroom with a child-friendly and trained person when the assessments were administered. Although participants missed small amounts of instructional time, a hallmark of excellent instruction is the use of assessments for instructional planning.
Progress monitoring is a current common practice in the schools where this study was conducted. The assessment procedures are familiar to the students as they already engage in this assessment within their school. As part of DIBELS Next and AIMSweb protocol, no criticism or corrections were made to children’s responses. Individual data were shared with the children’s teacher after the completion of the study. See Appendix K for an example of a graph that was provided to teachers. No results recorded and reported from this study include any identifying information regarding the child, teacher, or school. Possible presentations and publications will not use students’ real names. Pseudonyms or “subject” and descriptors such as gender and age may be used with no identifying school or personal names.

There were no risks to teachers related to participation in this study. The research took place over eight sessions within a time frame of less than 2 weeks. It would not be reasonable to expect detectable gains in student oral reading performance in this time period; therefore, while the research will allow evaluation of the assessment measures, the data cannot be used to measure instructional effectiveness in this time period.

Data were transported from the schools to Western Michigan University by the lead investigator and brought to a locked cabinet within the Special Education and Literacy Studies department in Dr. Kristal Ehrhardt’s office, 3406 Sangren. Data collected and organized in a computer file were kept on a password-protected computer, in a file that is further password-protected.
CHAPTER III

RESULTS

Overview

The primary objective of this study was to investigate further whether popularly used R-CBM passage sets are using well equated reading passages to monitor student reading growth. This study evaluated the homogeneity of oral reading fluency progress monitoring passages of two popular reading sets, DIBELS Next and AIMSweb. Currently, there is no agreed-upon method for determining the difficulty level of an R-CBM passage to equate reading passages. For teachers to make correct judgments about student academic progress, it is essential that R-CBM passages are of equal difficulty level. The purpose of this research is to examine the stability of each R-CBM progress monitoring passage set. This study investigated the variance within each passage set, as well as looked at how the variances compare between the two sets. Additionally, the passage sets were organized into triad groups for analysis and the variance within and between sets organized in this fashion was explored.

The data collected regarding the two progress monitoring passage sets were organized into four data sets for analysis. First, the 20 passages from DIBELS Next were examined as well as the 20 passages that were randomly selected from the 30 available AIMSweb progress monitoring passages. See Appendix H for specific information regarding which AIMSweb passages were included in the study. Additionally, DIBELS
Next has intentionally prearranged their 20 progress monitoring passage sets into a system of six triads and one dyad. The authors recommend that the passages be kept in this organized system when read by students and that those looking at the data should watch progress across the triads/dyad rather than at each individual progress monitoring assessment. The contention is that this will lead to better data to evaluate true reading progress. Powell-Smith et al. (2010) report that practitioners are “enhancing decisions about [student] skills and progress when a triad of passages is used” (p. 38). Therefore, the means of the triads and dyad organized by DIBELS Next were analyzed to investigate whether evaluating the student information in these organized triads provides information that has less variance associated with the difficulty of the passages read. AIMSweb, on the other hand, does not formally organize their passage sets into triads and trains that passages can be read in any order since they are intended to be of equal difficulty level. To investigate whether the purposeful organization of DIBELS Next is what has affected the variance, comparison with AIMSweb passages arbitrarily ordered into triads has occurred. AIMSweb passages were organized into triads based on no particular method and were not read in specific triads by the students. Therefore, this fourth data set is an arrangement of AIMSweb progress monitoring passages into triads (and one dyad).

Self-Assessments, Procedural Checklists, and Interrater Reliability

At each assessment session, those administering R-CBM passages reported their adherence to the procedural protocol. The Self-Assessment Checklist is included as Appendix I. Results indicate that the assessors reported that they adhered to the assessment procedures 98.3% of the time. Most commonly reported procedural errors
included issues with starting the timer at the correct time and with reading the standardized directions.

In an effort to analyze further the quality of the data taken by those administering the R-BCM probes for this study, interrater reliability was checked and a procedural checklist was conducted for 15% of the total assessment sessions. Results from the procedural checklists (Appendix J) indicate that 98.7% of the time the assessor held fast to the standard procedures. Interrater reliability, assessed through point-by-point agreement, was found to be at 99%. Total percentage of agreements was calculated for each probe by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (House et al., 1981).

**Variance Within and Between Passage Sets**

Due to the randomization and counterbalancing of the reading passages, practice and interaction effects are minimal and therefore allow for comparisons of variance, error, and overall means of the passages. The means for each passage are reported in Table 2. For the DIBELS Next passage set, the grand mean across all R-CBM passages was 117.8 wc/min. The lowest mean for any passage was 103.4 and the highest was 126.6. When investigating the AIMSweb passage set, the grand mean was 125.3 with the lowest mean for any passage 106.5 and the highest 138.7. Deviation scores were also calculated for each passage set to assist in determining variance across passages within the passage set and are also reported in Table 2.
<table>
<thead>
<tr>
<th>Passage</th>
<th>DIBELS Next</th>
<th>AIMSweb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Deviation Score</td>
</tr>
<tr>
<td>1</td>
<td>118.2</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>113.2</td>
<td>–4.7</td>
</tr>
<tr>
<td>3</td>
<td>126.6</td>
<td>8.8</td>
</tr>
<tr>
<td>4</td>
<td>123.0</td>
<td>5.2</td>
</tr>
<tr>
<td>5</td>
<td>111.2</td>
<td>–6.6</td>
</tr>
<tr>
<td>6</td>
<td>121.9</td>
<td>4.1</td>
</tr>
<tr>
<td>7</td>
<td>121.8</td>
<td>3.9</td>
</tr>
<tr>
<td>8</td>
<td>103.4</td>
<td>–14.4</td>
</tr>
<tr>
<td>9</td>
<td>125.4</td>
<td>7.6</td>
</tr>
<tr>
<td>10</td>
<td>115.0</td>
<td>–2.9</td>
</tr>
<tr>
<td>11</td>
<td>111.4</td>
<td>–6.4</td>
</tr>
<tr>
<td>12</td>
<td>126.4</td>
<td>8.5</td>
</tr>
<tr>
<td>13</td>
<td>116.2</td>
<td>–1.7</td>
</tr>
<tr>
<td>14</td>
<td>108.2</td>
<td>–9.6</td>
</tr>
<tr>
<td>15</td>
<td>128.0</td>
<td>10.2</td>
</tr>
<tr>
<td>16</td>
<td>118.9</td>
<td>1.1</td>
</tr>
<tr>
<td>17</td>
<td>112.7</td>
<td>–5.2</td>
</tr>
<tr>
<td>18</td>
<td>124.9</td>
<td>7.0</td>
</tr>
<tr>
<td>19</td>
<td>115.5</td>
<td>–2.4</td>
</tr>
<tr>
<td>20</td>
<td>115.0</td>
<td>–2.9</td>
</tr>
</tbody>
</table>
Mean Euclidean Distance (MED) was calculated for each passage and is reported in Table 2. “Euclidean Distance is the square root of the sum of squared difference between repeated measurements” (Christ & Ardoin, 2009, p. 58). MED was examined because this type of statistic is robust to small sample sizes and examines distances between passages at all points on the distribution. Mean Euclidean Distance is a measure of the dissimilarity of an item, or passage, to all other passages in the set. The higher the number, the more dissimilar the passage is on average. When creating an R-CBM passage set those passages with the lowest MED are selected for inclusion to develop a group of passages with the least amount of variance. When analyzing already created passage sets, as is being accomplished here, it is useful to look across MED information to determine which passages have greater variability within and across the passages of the passage set.

To be able to further visually inspect the variance within and between passage sets, the deviation scores were graphed and are displayed in Figure 1.

![Figure 1](image)

*Figure 1.* Deviation scores (wc/min) on the 20 passages read from DIBELS Next (top graph) and AIMSweb (bottom graph).
A repeated measures Analysis of Variance (ANOVA) was run to arrive at the estimated variance of each of the three distinct components: variation based on student reading results, variation based on error related to passage difficulty level, and random/unexplained error. Analysis occurred separately for DIBELS Next and AIMSweb passages sets as well as for each set organized into triads. The results, organized in Table 3, show where the variance is coming from for each set of passages. These data can be investigated for each passage set individually, and then also compared across sets to look at variation both within passage sets and across the sets. The passage set with lower item error is considered a more homogenous passage set.

Table 3

*Estimate of Variance Components*

<table>
<thead>
<tr>
<th>Sources of Variation</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>Estimated Variance Components</th>
<th>Percentage of Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DIBELS Next (non-Triad)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons (p)</td>
<td>660048.34</td>
<td>44</td>
<td>15001.10</td>
<td>743.53</td>
<td>81.0</td>
</tr>
<tr>
<td>Items (i)</td>
<td>39603.28</td>
<td>19</td>
<td>2084.38</td>
<td>43.42</td>
<td>4.7</td>
</tr>
<tr>
<td>Residual (pi,e)</td>
<td>109183.22</td>
<td>836</td>
<td>130.60</td>
<td>130.60</td>
<td>14.2</td>
</tr>
<tr>
<td><strong>AIMSweb (non-Triad)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons (p)</td>
<td>746704.28</td>
<td>44</td>
<td>16970.55</td>
<td>841.80</td>
<td>80.7</td>
</tr>
<tr>
<td>Items (i)</td>
<td>59975.60</td>
<td>19</td>
<td>3156.61</td>
<td>67.16</td>
<td>6.4</td>
</tr>
<tr>
<td>Residual (pi,e)</td>
<td>112513.50</td>
<td>836</td>
<td>134.59</td>
<td>134.59</td>
<td>12.9</td>
</tr>
<tr>
<td><strong>DIBELS Next Triads</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons (p)</td>
<td>228612.70</td>
<td>44</td>
<td>5195.74</td>
<td>733.72</td>
<td>92.2</td>
</tr>
<tr>
<td>Items (i)</td>
<td>1092.57</td>
<td>6</td>
<td>182.09</td>
<td>2.72</td>
<td>0.3</td>
</tr>
<tr>
<td>Residual (pi,e)</td>
<td>15753.70</td>
<td>264</td>
<td>59.67</td>
<td>59.67</td>
<td>7.5</td>
</tr>
<tr>
<td><strong>AIMSweb Triads</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons (p)</td>
<td>263361.41</td>
<td>44</td>
<td>5985.49</td>
<td>848.93</td>
<td>92.2</td>
</tr>
<tr>
<td>Items (i)</td>
<td>8052.20</td>
<td>6</td>
<td>1342.03</td>
<td>28.87</td>
<td>3.1</td>
</tr>
<tr>
<td>Residual (pi,e)</td>
<td>11337.53</td>
<td>264</td>
<td>42.95</td>
<td>42.95</td>
<td>4.7</td>
</tr>
</tbody>
</table>
Results show that, for each data set, the largest amount of variation was due to the student performance. Both the DIBELS Next and AIMSweb passage sets, when evaluated by individual passage rather than in triads, showed equal levels of variance, approximately 81%, due to student reading factors. Clearly, when the passages are analyzed across triads rather than through a point-by-point analysis, there is a decrease in variance related to error and an increase in variance related to student reading. In DIBELS Next, the authors recommend that R-CBM passages be administered and analyzed in triads. This is not common practice for AIMSweb.

Table 3 also shows the amount of error variance associated with the items versus error related to unaccounted sources. The amount of error variance associated with the items for both the DIBELS Next and AIMSweb passage sets when analyzed with independent passages shows very similar results with 6.4% of the error for AIMSweb and 4.7% of the error for DIBELS Next being related to passage difficulty. When the data are organized into the triad groups, both DIBELS Next and AIMSweb triads showed significant increase in variance related to student performance and decrease associated with error. Both data sets showed that 92.2% of the variance was related to student reading. DIBELS Next, which was purposefully organized into triads by the authors, did show less error associated with passage difficulty with less than 1% of the error connecting to this factor. For AIMSweb passages, which were not purposefully organized into triads and were not read in any particular order by the students in this study, about 3% of the error was associated with difficulty level of the passages read.
CHAPTER IV

DISCUSSION

For educators to be able to track accurately and plan regarding students’ progress in the area of reading, it is an essential component of R-CBM that the passages used for assessment are equivalent in difficulty level. The only way for teachers to know if student progress is due to true reading improvement is if passages are of similar difficulty level. If this criterion is not met, fluctuations in student performance may be due to variation in passage difficulty level, rather than genuine changes in student skill. For a passage set to be considered to have well developed readability, any difference in score should be due to student skill rather than due to differences in difficulty of the passage or other error.

Christ and Ardoin (2009) determined that developed passage sets using readability formulas have decreased variability when compared to passage sets created from a basal series, but they pointed out that, even with traditionally used readability formulas, much variance across passages remains. The results from this study indicate that, while using readability formulas to develop passage sets is an improvement, there continue to be high levels of error related to passage difficulty. The results of the current study show that, when analyzing student progress by each individual passage, approximately 81% of the variance in wc/min can be considered as due to student performance. This was true for both the DIBELS Next progress monitoring passage set and the AIMSweb passages.
Additionally, 4.7% of the error variance for DIBELS Next and 6.4% for AIMSweb were related to passage difficulty discrepancies.

DIBELS Next used a newly created readability formula, the DMG passage difficulty index, to equate the reading passages. The passage set continues to show that 4.7% of the error variance is related to continued readability inconsistencies across passages. However, this does show significant improvement from DIBELS version 6.0, which had 10% of the variance of the passage set related to passage difficulty (Poncey et al., 2005). While it may be that the newly developed readability formula played a role in developing passages sets that are of more equal reading level, a variety of procedural changes were implemented in the creation of this new passage set, including field testing of passages. DIBELS Next designers administered a number of produced passages to a trial group of students and then discarded passages that showed significant variability. This process was not done with DIBELS version 6.0, and Ardoin, Suldo, et al. (2005) state that this is critical step in creating a well developed R-CBM passage set.

Analyzing Data in Triads

The authors of DIBELS Next have purposefully organized their passage sets into triads and encourage those who are looking at student data to use the triads to investigate student progress. Powell-Smith et al. (2010) stated that using the prepared triads improves assessor judgment regarding true student progress. This current study, in agreement with the DIBELS Next authors, indicates that when the passages are organized into triads to examine student reading variance due to difficult level of the passages greatly decreases. In this instance, error related to item variance decreased from 4.7% when looking at each
passage reading individually to 0.3% when looking across arranged triads. Approximately 92% of the variance was related to student achievement rather than either item or random error, an increase of 11% from when passages are looked at one-by-one.

Interestingly, when the AIMSweb passage set, which is not intentionally organized into triads, was put into arbitrary triads, the amount of variance related to student performance decreased as well. Although AIMSweb has not purposefully organized their passages into groups of three for assessment and analysis, there is a benefit to the overall interpretation of the progress monitoring data if the passages are analyzed this way. Variance related to student performance increased to approximately 92% as well, and error related to passage difficulty decreased from 6.4% when investigating each passage individually to 3.1% when using arbitrary triads. Thus, whether the DIBELS Next set, which is intentionally organized into triads, is being used to progress monitor students, or another system without such intentional organization is utilized, teachers will have a better understanding of true student progress if they look across multiple progress monitoring passages/sessions to determine reading growth, rather than making an improvement determination at each individual assessment session. When the data are analyzed across three sessions, rather than one assessment, error related to item difficulty and random error are both significantly decreased and information better reflects true student progress.

It is important to note that there was less error related to difficulty level of the passages with the purposefully organized DIBELS Next passages. DIBELS Next purposeful organization into triads was able to decrease error related to passage difficulty to an exceptionally low level of error variance that has not yet been seen in the R-CBM
literature. It is evident that analysis regarding student progress can be improved regardless of what system of R-CBM is used if decisions are made after multiple assessments. Nevertheless, it is also apparent that the organization of DIBELS Next triads has strong results and it would benefit practitioners and researchers to consider the quality of the passage set when determining which R-CBM passages to use to monitor student reading growth.

Effect of Readability Formulas on the Interpretation of R-CBM Data

Ardoin, Suldo, Witt, Aldrich, and McDonald (2005) investigated popular readability formulas to determine whether there is a formula, or a component of a formula, that highly correlates with student reading and therefore should be used in developing R-CBM passages. They concluded that there are components of a reading passage that were found to correlate with student results, including number of syllables per 100 words, and number of words on the Dale-Chall 3,000 word list. The DMG Passage Difficulty Index attempted to use the information from this study, as well as others investigating readability, to develop an improved readability formula. From this study it is not clear that the newly developed formula on its own is able to provide a passage set with considerably increased variance related to student performance. Much like the AIMSweb passage set, when looking across all developed passages, 81% of the DIBELS Next passage variance continued to be related to student reading, and the rest related to error variance. Continued investigation of readability formulas and systems for developing R-CBM passage sets is warranted.
Differences in the Grand Means of DIBELS Next and AIMSweb

When considering how the two passages sets compare, it is interesting to also note the difference in grand means between the two passage sets. A variety of research has been done regarding what level of student oral reading fluency is necessary to be considered benchmark, or on track, with overall reading. Due to the difference in grand means between the two passage sets with this group of students, DIBELS Next showing a grand mean of 117.8 and AIMSweb showing 125.8, it is critical to consider whether cut scores, or benchmark levels, are able to be generalized across passage sets. This information would indicate that using a cut score based on one set of passages does not relate well to another passage set because the resulting student scores do seem to vary systematically based on passage development and resulting error variances related to each individual passage set.

Limitations

While the results presented from this research provide important information to those who use R-CBM for progress monitoring of student reading performance, several limitations of the current study should be addressed. Students included in this study were from one school district, largely Caucasian, in the Midwest. The limitations of the diversity within the population assessed may decrease the ability to generalize these results. Variance based on passage difficulty may be larger or smaller in a different area of the country, or with students of different backgrounds and experiences. It may be beneficial in the future to investigate a more diverse group of participants.
Furthermore, this study was conducted with third grade students, and with only third grade passage sets from DIBELS Next and AIMSweb. Other grade levels may have different results. Variance in reading R-CBM passages in first grade, or in sixth grade, may not follow the patterns and results discussed here. Investigation of the variance related to passage difficulty at other grade levels may be warranted so that those who interpret R-CBM information may have an enhanced understanding of the best way to use and decipher student data at each grade level. Also, this study looked at students reading in the strategic and benchmark ranges due to grade level reading goals being appropriate for students reading at this level. It may not be appropriate to generalize results to students who are in the intensive range and are being progress monitored with either grade level, or out of grade level, materials.

Assessing students with multiple probes across 8 sessions is not typical progress monitoring procedure. Students are typically assessed with one passage weekly, biweekly, or monthly, depending on the teacher preference. Then, student progress is monitored across time and slope is inspected. This study was conducted with five reading passages presented to the student at each session, and sessions were completed in a very short period of time, across a week and a half. While this method assisted in looking closer at homogeneity of passages, a study investigating how homogeneity holds up and what type of variance is found across time would be beneficial. As Fuchs (2004) stated, this type of research is less common due to the logistics involved in assessing a student across the length of time needed for full analysis.
Conclusion

The current study supports the authors of DIBELS Next in their organization of the progress monitoring passages into groups for analysis, rather than making decisions regarding student progress after each assessment session. Through the analysis of the AIMSweb passages, it is clear that any progress monitoring interpretation is enhanced when progress is watched across multiple measures. The commitment of those conducting student assessments and analyzing achievement is to use accurate academic measurement, with the least amount of error in the assessment. When utilizing any form of R-CBM, if educators utilize the practice of waiting to collect at least three further progress monitoring data points prior to making a decision about true progress they will be using a procedure that will significantly decrease error and increase how much observed variation in score is truly based on student reading performance.
REFERENCES


Appendix A

R-CBM Directions and Scoring from the DIBELS Next Manual
Administration Directions

Follow these directions exactly each time with each student. Say the words in bold italic type verbatim. Put the student copy of the materials in front of the student and say the following:

I would like you to read a story to me. Please do your best reading. If you do not know a word, I will read the word for you. Keep reading until I say “stop.” Be ready to tell me all about the story when you finish. (Place the passage in front of the student.)

Begin testing. Put your finger under the first word (point to the first word of the passage). Ready, begin.

1. Do not read the title to the student. If the student chooses to read the title, do not start the stopwatch until he/she reads the first word of the passage. If the student asks you to tell him/her a word in the title or struggles with a word in the title for 3 seconds, say the word. Do not correct any errors the student makes while reading the title.

2. Start the stopwatch after the student says the first word of the passage. If the student is silent or struggles for 3 seconds with the first word of the passage, mark the word as incorrect, say the word, and start the stopwatch.

3. During benchmark assessment, three passages are administered if the student reads 10 or more words correctly on the first passage. When administering the second and third passages, use the following shortened directions:

Now read this story to me. Please do your best reading. Ready. Begin.

4. During the testing:
   - Follow along in the scoring booklet.
   - Leave blank any words read correctly. Mark a slash (/) through errors (including skipped words).
   - The maximum wait time for each word is 3 seconds. If the student does not provide the word within 3 seconds, say the word and mark it as incorrect.
   - During benchmark assessment, students read 3 different passages, for 1 minute each. If the student reads fewer than 10 words correctly on the first passage, record his/her score for words correct and errors on the front cover of the booklet and do not administer passages 2 and 3.
   - At the end of 1 minute, place a bracket ([ ]) in the text after the last word provided by the student. Say Stop and remove the passage. If the student completes the assessment before 1 minute, stop testing and record the student’s score. Scores are not prorated.

Note: If the student is in the middle of a sentence at the end of 1 minute, you may allow the student to finish the sentence, but only score the words said up to the end of 1 minute.

5. If the student reads 40 or more words correctly on the passage, have the student retell what he/she has just read using the directions provided below. If the student reads fewer than 40 words correctly on a passage, use professional judgment whether to administer Retell for that passage.
Immediately after testing:

- Score reading passages immediately after administration. Use the cumulative word count to determine the total words read. Record the total words on the Total Words line on the scoring sheet.
- Record the number of errors (including skipped words) on the Error line on the scoring sheet for each passage.
- Subtract the number of errors from the total words to get the number of words correct and record it on the Words correct line.

**Scoring Rules for DORF**

*The student receives 1 point for each word read correctly in 1 minute.*

1. Leave blank any words the student reads correctly. Inserted words are not counted. To be counted as correct, words must be read as whole words and pronounced correctly for the context of the sentence.

2. Mark a slash (/) through any errors. Errors include words read incorrectly, substitutions, skipped words, hesitations of more than 3 seconds, words read out of order, and words that are sounded out but not read as a whole word.

**Discontinue Rule**

Discontinue administering DORF if the student reads zero words correctly in the first line of the first passage. Record a score of 0 on the Total line on the scoring page and in the DORF score box on the cover page of the student booklet. If the student reads fewer than 10 words correctly on the first passage during benchmark assessment, do not administer Retell or the second and third passages. If the student reads fewer than 40 words correctly on any passage, use professional judgment on whether to administer Retell for that passage.

**Wait Rule**

Wait 3 seconds for the student to respond. If the student hesitates for 3 seconds on a word, mark a slash (/) through it and read the word to the student. If necessary, indicate for the student to continue with the next word by pointing.

**Reminders**

If the student stops reading (and it’s not a hesitation on a specific item), say *Keep going. This reminder may be used as often as needed.*

If the student loses her/his place while reading, point. *This reminder may be used as often as needed.*

**Note:**

Students are not penalized for differences in pronunciation due to dialect, articulation delays or impairments, or for pronunciations due to speaking a first language other than English.
Appendix B

Human Subjects Institutional Review Board
Letter of Approval
Date: February 16, 2011

To: Kristal Ehrhardt, Principal Investigator
   Christine Russell, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 11-02-24

This letter will serve as confirmation that your research project titled “Making Valid Instructional Decisions Based on Progress Monitoring with DIBELS Next and AMIS web: Estimates of Standard Error of Measurement” has been approved under the expedited category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: February 16, 2012
Appendix C

Teacher Consent
Your school/classroom has been invited to participate in a study titled “Making Valid Instructional Decisions Based on Progress Monitoring with DIBELS Next and AIMSweb: Estimates of Standard Error of Measurement.” This consent document will explain the purpose of this study and will go over all of the time commitments, the procedures used in the study, and the risks and benefits to the school, staff, students, and educational community, of participating in this project.

**What are we trying to find out in this study?**
The purpose of this project is to evaluate the consistency of the readability level of passages of DIBELS Next and AIMSweb. As you know, these assessments are used to progress monitor student reading growth. To have an accurate picture of student growth, the passages used must be at a similar readability level. The goal of this research is to examine the stability of each R-CBM progress monitoring passage set to determine whether schools are using a well developed set of passages to progress monitor their student’s reading achievement.

**Who can participate in this study?**
Participants in the study are third grade students identified, through the winter DIBELS Next benchmark assessment that the school has already conducted, as reading in the benchmark or strategic range.

**Where will this study take place?**
The testing will occur in an assigned testing space in the school building. So that students are most comfortable, testing will occur in the classroom, in the hallway outside the classroom, or in an area adjoining the classroom. Location will be established in consultation with the children’s classroom teacher and the building principal.

**What is the time commitment for participating in this study?**
The assessment process for each child will occur across eight testing sessions, all completed in a two week window, where 5 one-minute times reading passages will be read by each student. The testing will occur in March, 2011.

**What will you be asked to do if you choose to participate in this study?**
You as the classroom teacher will be asked to help identify a time of day when it is appropriate to take students out of the classroom for the assessment. Also, teachers will meet with the student investigator prior to beginning the project for a planning meeting, and will assist in identifying students who meet the criteria for involvement in the study.
Students will leave the classroom for eight testing sessions across eight days. They will be asked to read 5 one-minute timed reading passages at each session. This should take approximately eight minutes per student to complete per day.

**What information is being measured during the study?**
This study is measuring DIBELS Next and AIMSweb third grade progress monitoring passage set homogeneity. Students will read the passages and the words read correct for each passage will be recorded.

**What are the risks of participating in this study and how will these risks be minimized?**
There are no more than minimal risks for student participation in this study, and these risks do not exceed standard risks involved when school officials conduct their R-CBM assessments. Students will leave their classroom with a child-friendly and trained person when the assessments are administered. Although potential participants will miss small bits of instructional time, a hallmark of excellent instruction is the use of assessments for instructional planning. In addition, the student investigator will work closely with school personnel to be sure the assessments occur at a time in the day when neither critical instruction nor important activities such as lunch time are missed.

Progress monitoring is a current common practice in the schools that this study will be conducted in. The assessment procedures are familiar to the students as they already engage in this assessment within their school. As part of DIBELS Next and AIMSweb protocol, no criticism or corrections are made to children’s responses. Individual data will be shared with the children’s teacher, but results reported in the study will not include any identifying information regarding the child, teacher or school.

There are no risks to teachers related to participation in this study. The research will take place over 8 sessions within a time frame of less than two weeks. It would not be reasonable to expect detectable gains in student oral reading performance in this time period; therefore, while the research will allow us to evaluate these assessment measures, the data cannot be used to measure instructional effectiveness in this time period.

**What are the benefits of participating in this study?**
This data will investigate how much variance is associated with each commonly used R-CBM progress monitoring passages set and the data will be used to assist in determining the quality of these commonly used assessments. This research is intended to benefit the educational community in gaining additional information regarding the use of these assessments for evaluation of student reading achievement and progress. Further, Parchment school administrators have adopted these progress monitoring procedures. The data will help the school evaluate the use of these assessments in their own school.

**Are there any costs associated with participating in this study?**
There are no costs associated with participating in the project.

**Is there any compensation for participating in this study?**
For participating in the study, each teacher will receive a $50 gift card from the Teacher’s Center in Portage, Michigan, to purchase classroom materials. Also, the student investigator is trained in a variety of educational professional development areas and would like to, as a thank you to the school personnel, provide a free professional development seminar for each school in the area of
DIBELS Next, data analysis, or reading instruction. This will be scheduled during the 2011-2012 school year if the school desires such professional development.

**Who will have access to the information collected during this study?**
The student investigator will keep all data and information collected in a password protected file on her personal computer. During the study, when assessments are not being completed, all student booklets and folders with data will be kept in a secure place in the elementary school’s office. Upon completion of the research, data that is kept will have no student, teacher, or school identifying information and will be kept in a locked cabinet at Western Michigan University in the Department of Special Education and Literacy Studies.

**What if you want to stop participating in this study?**
Students, teachers, or principals can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either professionally or personally if you choose to withdraw from this study.

Should you have any questions prior to or during the study, you can contact the primary investigator, Dr. Kristal Ehrhardt at (269) 387-4478 or kristal.ehrhardt@wmich.edu or the student investigator, Christine Russell at (616) 318-7111 or christine.r.russell@wmich.edu. You may also contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions or problems arise during the course of the study.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

Please Print Your Name

__________________________________________  ______________________
Teacher’s signature              Date
Appendix D

Parent Notification
Western Michigan University

Department of Special Education and Literacy Studies

Principal Investigator: Kristal Ehrhardt, Ph.D.
Student Investigator: Christine Russell, Ed.S.
Title of Study: Making Valid Instructional Decisions Based on Progress Monitoring with DIBELS Next and AIMSweb

Dear Parent(s),
My name is Christine Russell and I am conducting a research study in your child’s class.

WHAT ARE WE STUDYING?
I am looking at the accuracy and usefulness of one-minute timed reading passages that are used at Parchment School District to look at student’s reading.

WHY?
This project will serve as my dissertation for the requirement of a Doctor of Education. Also, we are looking to further the educational community’s understanding of these one-minute timed reading tests.

HOW DOES THIS AFFECT ME AND MY CHILD?
I, or a research assistant, will work with your child in March, 2011 for 8 sessions. Every time we work with your child they will read 5 one-minute timed reading passages. Each time they are out of the classroom it will be for about eight minutes. I will work closely with your child’s teacher to be sure it is not during important instruction time. Information that is kept for this study will not include your child’s name.

WHAT BENEFITS WOULD MY CHILD GET FROM THIS?
Most children find these test fun. We could get some information that could help your child’s teacher with improving instruction for your child. If your child seems uncomfortable or unhappy, we will stop the assessments.

HOW CAN I GET MORE INFORMATION?
If you have any questions or concerns about the study, or if you would like to withdraw your child from the study, please contact Christine Russell at (989) 670-2411, Dr. Kristal Ehrhardt at (269) 387-4478, your child’s teacher, or your child’s principal.
Appendix E

Student Assent Form
Western Michigan University
Department of Special Education and Literacy Studies

Principal Investigator: Kristal Ehrhardt, Ph.D.
Student Investigator: Christine Russell, Ed.S.
Title of Study: Making Valid Instructional Decisions Based on Progress Monitoring with DIBELS Next and AIMSweb: Estimates of Standard Error of Measurement

*The Student Investigator or a Research Assist will orally read the following to the student and after reading will ask the student to write their name in the blank if they want to be a part of the study:*

We are doing a research study. A research study is a special way to find out about something. We want to find out how good the reading passages are that your teacher uses to test your reading progress.

If you want to be in this study, you will be asked to read with one of us every day for 8 days. When you read with us, you will read 5 one-minute reading passages.

If you do this reading with us, you may miss a small bit of class, but your teacher knows what you are doing and will help to be sure you get caught back up.

If you decide to be in this study, you will be helping us make sure that we’re using good reading passages and that we are using good information to make decisions about your reading.

When we are done with the study, we will write a report about what we found out. We won’t use your name in the report.

You don’t have to be in this study. You can say “no” and nothing bad will happen. If you say “yes” now, but you want to stop later, that’s okay too. No one will be mad at you, or punish you if you want to stop. All you have to do is tell us you want to stop.

If you have any questions or concerns about this study, you can ask your teacher or any of us that you read with.

If you want to be in this study, please write your name.

I, ____________________________________, want to be in this research study.

(write your name here)

_____________________________________   ___________   ___________
Researcher’s Signature       (Date)
Appendix F

Example of Student Materials for an R-CBM Probe
Learning to Skateboard

The last box was finally unpacked at the new house. Zach had been excited about moving to the city and about making new friends. He knew he had several weeks to explore the neighborhood before school started, and he realized that was plenty of time to learn his way around. He asked his mom if he could walk to the nearby park.

From the road, the park looked like a giant green expanse. Now, he saw that it was divided into different sections. The first thing that caught his eye was the skate park. He sprinted to the gate and stood looking in at the ramps and rails. He had a skateboard but had never learned to ride it. The roads near his old house had been too rocky.

Just then, two boys zoomed up on skateboards and said hello. They asked him if he skateboarded. He told them that he had a board but had never had a place to learn. “Well, now you do,” they said, and they offered to teach him. Their names were Matt and Pablo, and they went to the same school he was going to attend. Zach happily agreed to meet them for a lesson.

The following afternoon, Zach went to find the boys. After putting on helmets and pads, Matt showed him how to stand on the board above the axle. Pablo helped him practice pushing off, which is the movement needed to get the board rolling. They also explained carving and stopping. Zach practiced and also watched the other guys perform some advanced moves they had mastered. After thanking his new friends, he promised to come back every day. By the time school started, he had learned to skateboard and had made two good friends.
Appendix G

Example Scored R-CBM Reading Probe
DIBELS® Oral Reading Fluency
Level 3/Progress Monitoring 1

Total words: 98
Errors (include skipped words): 6
Words correct: 92

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.
Appendix H

AIMSweb Passages Randomly Selected for This Study
AIMSweb Passages Randomly Selected for this Study

<table>
<thead>
<tr>
<th></th>
<th>AIMSweb Third Grade Passage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>20</td>
<td>33</td>
</tr>
</tbody>
</table>
Appendix I

R-CBM Self-Assessment Accuracy Checklist
R-CBM Self-Assessment Accuracy Checklist

*Modified from the DORF Assessment Accuracy Checklist (Good and Kaminski, 2010) and the AIMSweb Accuracy of Implementation Rating Scale (Shinn, 2004)*

<table>
<thead>
<tr>
<th>Followed Protocol</th>
<th>Did Not Follow Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Position materials so that student cannot see what is being recorded?</td>
<td></td>
</tr>
<tr>
<td>2. Seated appropriate distance from reader?</td>
<td></td>
</tr>
<tr>
<td>3. State standardized directions exactly as written?</td>
<td></td>
</tr>
<tr>
<td>4. Start the timer when the student reads the first word of the passage?</td>
<td></td>
</tr>
<tr>
<td>5. Score student responses correctly according to the scoring rules?</td>
<td></td>
</tr>
<tr>
<td>6. Use reminder procedures correctly and appropriately?</td>
<td></td>
</tr>
<tr>
<td>7. Say the word and put a slash over it if the student fails to say it correctly within 3 seconds</td>
<td></td>
</tr>
<tr>
<td>8. Write “sc” above a previously slashed word if the student self-corrects within 3 seconds?</td>
<td></td>
</tr>
<tr>
<td>9. Discontinue if the student does not read any words correctly in the first row of the passage?</td>
<td></td>
</tr>
<tr>
<td>10. Place a bracket (</td>
<td>) after the last word the student read before the minute ran out and tell the student to stop?</td>
</tr>
<tr>
<td>11. Correctly calculate the total number of words read (correct and errors) and record it on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>12. Correctly add the number of errors and record it on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>13. Correctly subtract the errors from the total words and record the words correct on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>14. Gave correct passage to the student?</td>
<td></td>
</tr>
<tr>
<td>15. Gave passages in the correct order to the student?</td>
<td></td>
</tr>
</tbody>
</table>
Appendix J

R-CBM Procedural Integrity Assessment Accuracy Checklist
# R-CBM Procedural Integrity Assessment Accuracy Checklist

*Modified from the DORF Assessment Accuracy Checklist (Good & Kaminski, 2010) and the AIMSweb Accuracy of Implementation Rating Scale (Shinn, 2004)*

<table>
<thead>
<tr>
<th>Followed Protocol</th>
<th>Did not Follow Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Position materials so that student cannot see what is being recorded?</td>
<td></td>
</tr>
<tr>
<td>2. Seated appropriate distance from reader?</td>
<td></td>
</tr>
<tr>
<td>3. State standardized directions exactly as written?</td>
<td></td>
</tr>
<tr>
<td>4. Start the timer when the student reads the first word of the passage?</td>
<td></td>
</tr>
<tr>
<td>5. Score student responses correctly according to the scoring rules?</td>
<td></td>
</tr>
<tr>
<td>6. Use reminder procedures correctly and appropriately?</td>
<td></td>
</tr>
<tr>
<td>7. Say the word and put a slash over it if the student fails to say it correctly within 3 seconds</td>
<td></td>
</tr>
<tr>
<td>8. Write “sc” above a previously slashed word if the student self-corrects within 3 seconds?</td>
<td></td>
</tr>
<tr>
<td>9. Discontinue if the student does not read any words correctly in the first row of the passage?</td>
<td></td>
</tr>
<tr>
<td>10. Place a bracket (J) after the last word the student read before the minute ran out and tell the student to stop?</td>
<td></td>
</tr>
<tr>
<td>11. Correctly calculate the total number of words read (correct and errors) and record it on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>12. Correctly add the number of errors and record it on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>13. Correctly subtract the errors from the total words and record the words correct on the scoring sheet?</td>
<td></td>
</tr>
<tr>
<td>14. Gave correct passage to the student?</td>
<td></td>
</tr>
<tr>
<td>15. Gave passages in the correct order to the student?</td>
<td></td>
</tr>
</tbody>
</table>
Appendix K

Example Teacher Report of Study Data
Example Teacher Report of Study Data

*At the completion of the study, teachers were furnished with a graph for each individual participating student depicting their performance across all reading passages from DIBELS Next and AIMSweb.