Scholar Commons
USF University of South Florida

University of South Florida
Scholar Commons

Graduate Theses and Dissertations | Graduate School

2004

Assessing student reading progress: A comparison of generic and basal curriculum-based reading probes

Kelli S. Henson
University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/etd

 Part of the American Studies Commons

Scholar Commons Citation
Henson, Kelli S., "Assessing student reading progress: A comparison of generic and basal curriculum-based reading probes" (2004). *Graduate Theses and Dissertations.*
http://scholarcommons.usf.edu/etd/1075

This Ed. Specalist is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Assessing Student Reading Progress: A Comparison of Generic and Basal Curriculum-Based Reading Probes

by

Kelli S. Henson

A thesis submitted in partial fulfillment
of the requirements for the degree of
Education Specialist
Department of Psychological and Social Foundations
College of Education
University of South Florida

Major Professor: Kelly A. Powell-Smith, Ph.D.
Kathy L. Bradley-Klug, Ph.D.
Robert F. Dedrick, Ph.D.

Date of Approval:
November 30, 2004

Keywords: curriculum-based measurement, assessment, fluency, literacy

© Copyright 2005, Kelli S. Henson

Table of Contents

List of Tables

List of Figures

Assessing Student Reading Progress: A Comparison of Generic and Basal Curriculum-Based Reading Probes

Kelli S. Henson

ABSTRACT

The purpose of this study was to clarify and extend previous research on the comparibility of curriculum-based measurement oral reading fluency results using reading materials from outside of the students' curriculum for repeated measurement over time. Specifically, this study evaluated the use of generic measurement materials for monitoring student reading growth and expected gains in words read correctly per minute over time at different grade levels. Sixty-four first through third grade students were assessed twice weekly using both AimsWeb and Open Court reading probes. The dependent variables in this study were the level, which is defined as the mean of the data points for each type of probe, and the slopes derived from the number of words read correctly across all of the data collection days. A 3 (grade) x 2 (probe type) repeated measures ANOVA using the three grade levels as a between group variable and the two probe types as a within group variable was conducted with slopes as the dependent measure as well as with level as the dependent measure.

Analysis of levels revealed a significant ($p<.05$) main effect for probe type with significantly higher levels found in AimsWeb probes when compared to Open Court probes. There was also a significant ($p<.05$) main effect for grade with WRCPM

increasing with increasing grade level. However post hoc analysis revealed that the level difference was significant between first and third grade only.

The slope analysis revealed a significant main effect for grade with students in first and second grades making more progress than students in third grade. The slopes were higher for first and third graders in AimsWeb, but higher for second graders in Open Court probes. Slopes were not significantly different based on the type of probe was used to monitor progress. Both AimsWeb Open Court probes displayed sensitivity in the measurement of reading growth over time.

This study provided support for the use of measurement materials from outside the curriculum for CBM progress monitoring. Specifically, generic, noncurriculum-based probes were equally as sensitive to growth over time as curriculum specific materials.

Chapter One

Introduction

Reading has become a focus for much recent research and legislation (e.g, No Child Left Behind). A recent National Adult Literacy Survey reported that about 50% of adults who participated were not able to read at a proficient level. In addition, adults with "proficient" reading skills were significantly more likely to graduate from high school than less proficient peers (Sum, 1999). The National Reading Panel (2000) reported that over forty billion dollars is spent each year by US companies due to illiteracy.

Monitoring student reading progress has become an important focus of research in response to the need for increased accountability for student outcomes. A promising way to investigate student reading performance and growth is through assessment of reading fluency. Research has established measures of oral reading fluency as valid and reliable predictors of both reading outcomes and performance on high stakes tests (Good, Simmons, & Kame'enui, 2001). The National Reading Panel, which was convened to assess the research on strategies for reading instruction of children, identified reading fluency as an important topic for research in reading. Reading fluency can be defined as the "rate and accuracy in oral reading" (Shinn, Good, Knutson, Tilly, & Collins, 1992). Reading fluency has also been shown to correlate highly with comprehension (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988), another of the important

reading research topics identified by the National Reading Panel. Reading fluency bridges the gap between recognition and comprehension.

Curriculum-Based Measurement

Curriculum-based measurement is a vehicle by which reading fluency can be frequently measured in order to assess student progress in reading. Curriculum-based measurement (CBM) is a system of measuring student performance with standardized measurement procedures within the curriculum used in the classroom. Data from CBM in reading are used to make educational decisions for referrals, developing Individual Education Plans, program development, and program evaluation (Deno, 1985). CBM in reading is used in both general education and special education settings to assess student reading progress and early identification of students at-risk for reading failure. Additionally, measurements of student outcomes can be used to assess the success of educational programs and CBM provides for frequent assessment of student outcomes. CBM is easy to administer and efficient to allow for repeated measurement. It is clear, easy to understand, and economical. CBM is also standardized and has established reliability and validity (Deno, 1985).

Development and Analysis of Monitoring Probes

Traditional CBM of reading fluency requires reading probes for progress monitoring to be developed from the curriculum in which students are currently being instructed (Deno, 1985). Reading materials for monitoring reading progress are randomly sampled from the students' curriculum and passages are typed on separate paper with no pictures. These reading selections are called reading probes. Recent literature suggests the possibility that reading probes may not need to be developed from

the curriculum in which the student is taught to accurately show growth over time. Despite some contradictory findings (Hintze, Shapiro, & Lutz, 1994), some support has been shown for the claim that CBM oral reading passages developed from material outside students' curriculum are sensitive to growth over time. Oral reading probes developed from outside students' curriculum have been able to show comparable growth over time to probes developed from curriculum materials, but the number of words read correctly have been shown to vary according to probe type (Hintze & Shapiro, 1997; Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; Powell-Smith & Bradley-Klug, 2001).

In 1999, Edformation began marketing AIMSweb probes, generic measurement materials for monitoring reading. Whereas teachers or other school staff would have to develop probes, record and interpret information, and graph results using traditional methods, Edformation provides support for schools and parents who use the probes by providing reading probes, organizing progress monitoring results, and providing graphical representations of student progress. Additionally, the reading difficulty of traditional reading probes tend to vary from probe to probe within each grade level (Fuchs & Deno, 1994). However, AIMSweb probes are closely controlled for readability because they are developed outside of a specific curriculum (Howe & Shinn, 2001).

The use of generic probes, probes not developed from a specific curriculum, seems promising for convenience and consistent readability, but little research has been conducted using CBM reading probes that are not developed from a specific curriculum. Powell-Smith and Bradley-Klug (2001) compared basal probes to Tests of Reading Fluency (TORF) probes that are generic. Results showed that the TORF probes and the

basal probes were equally sensitive measures of reading growth over time. However, this type of research has not been done using the AIMSweb probes.

Expected Reading Growth Rates

Several researchers have investigated reading growth standards and expected oral reading growth rates. However, the outcomes of this research have varied (Deno, Fuchs, Marston, & Shin, 2001; Marston & Magnusson, 1985). In addition, researchers have not used generic progress monitoring reading materials in their studies. Although King, Bradely-Klug, Knoff, and Powell-Smith (2002) investigated the effect of different types of materials used for progress monitoring on expected growth rates for special education students, this comparison has not been made in general education.

The focus of this study is on determining if probes drawn from students' curriculum material and from generic material are equally as sensitive to student reading growth over time and the number of words read correctly per minute. In addition, expected reading growth rates are investigated as a function of both grade level and type of probe.

Participants for this study were students in first through third grade enrolled at a central Florida charter school for students who have frequent out-of-school suspensions. The charter school defines out-of-school suspension by the Florida State Department of Education (1995) guidelines as "the removal of a student from the public school environment for a period not in excess of 10 days." During the study, the participants' reading progress was monitored twice a week over a 20-week period using probes from the students' curriculum and probes developed outside of the curriculum. Each teacher conducted progress monitoring and data collection in her individual classroom.

Purpose

The purpose of the current study is to clarify and extend previous research on the comparibility of CBM oral reading fluency (ORF) results using reading materials from outside of the students' curriculum for repeated measurement over time. Specifically, this study evaluated the use of generic measurement materials for progress monitoring student reading growth. Additionally, this study investigated expected gains in words read correctly per minute over time at different grade levels. The following questions were specifically addressed in this study:

1) For elementary students, do curriculum-based probes result in the same number of words read correctly per minute as standardized probes developed outside of a curriculum?

2) Do curriculum-based probes result in the same rate of increase over time in words read correctly per minute as standardized probed developed outside of a curriculum?

3) Does the rate of increase over time in words read correctly per minute differ at different grade levels?

4) Is there an interaction effect of type of probe and grade level for the rate of increase in words read correctly per minute over time?

The independent variables for this study were (a) type of material used for progress monitoring (AimsWeb probes from no specific curriculum, Open Court probes developed from basal readers), and (b) student grade level (first, second, third). The dependent variables will be (a) the mean of the number of words read correctly by each

participant for each type of CBM probe (level), and (b) the slopes derived from the number of words read correctly in each CBM probe type across time.

This study offers three main contributions to the present CBM ORF literature base. First, this study investigated the comparability of two sources of measurement material (i.e., AIMSweb, Open Court) for CBM ORF progress monitoring that have not been previously examined. Secondly, by collecting progress monitoring data for a longer period of time (i.e., 19 weeks) than other studies examining the relationship between reading growth over time and the material used to measure it (i.e., five to ten weeks), this study displayed a more complete picture of reading growth over time. The results of this research may help to clear up the disparity in findings. Third, by examining differences in expected growth rates due to differences in measurement materials, this study adds to the research on academic growth standards for students and clarifies reasons for differences in the results of previous research. In addition, this research has a practical impact in that it may inform decisions about the choice of progress monitoring materials and assist in the development of appropriate goals for improvement in ORF with the establishment of academic growth standards in varying curricula.

Chapter 2

Review of Literature

This chapter introduces the history and development of curriculum-based measurement (CBM) with concentration on the oral reading fluency (ORF) measure. This chapter includes a description of the technical features of the CBM ORF and expectations for student growth in reading. Specific research studies relating to the effects of the development of CBM reading materials from varying curriculum are presented. A discussion of the purpose of the current study concludes this chapter.

History and Development of CBM

Curriculum-based measurement (CBM) began as part of a research study at the University of Minnesota in the 1970's and has evolved over the past 23 years. The research was undertaken in an effort to help teachers to individualize and modify instruction based on measurement and evaluation procedures. CBM was seen as a way for teachers to monitor student progress within the curriculum (Deno, 1985). Thus bridging the gap between measurement and instruction. Traditional, published, and norm-referenced tests have been criticized for their inability to inform decision–making and support effective instruction. Curriculum-based measurement was developed as an alternative to these assessment tools (Fuchs & Deno, 1994).

Curriculum-based measurement was designed around four basic characteristics: (1) measures should have established reliability and validity, (2) must be easy to

administer and efficient to allow for repeated measurement, (3) should be clear and easy to understand, and (4) economical (Deno, 1985). The use of CBM in schools has several advantages including sensitivity to the rate of performance, not just performance accuracy (Tindal, 1988). CBM techniques are useful in instructional decision-making and sensitive to growth over short periods of time (Deno, 1985). Also, CBM techniques allow for improved communication between decision-makers and the ability to identify student difficulty and adjust instruction in a timely manner. CBM aligns the curriculum with the assessment of student skills and knowledge by allowing teachers to directly verify the extent to which learning has occurred (Fuchs & Deno, 1994). With CBM, students' skills can be readily compared to those of their peers and measurement can be repeated over time (Deno, 1985).

Shinn (1998, 2002) describes the current view of CBM in terms of three "big ideas." First, CBM measures are "dynamic indicators of basic skills" or DIBS. CBM is dynamic in its sensitivity to differences both among individuals and within an individual over time. CBM is an indicator in that it is based upon empirical data and is a key correlate of important academic behaviors and skills. CBM is designed to evaluate student's basic skills, not their performance in specific content areas. Shinn's second "big idea" is formative evaluation as the main purpose of CBM. CBM is designed for progress monitoring and evaluation continuously throughout a program, not just at the end of the program. Shinn's third big idea involves using CBM to make various decisions as part of a problem-solving model. A problem solving model usually consists of approximately five decision stages including problem identification, problem

certification, exploring solutions, evaluating solutions, and problem solution (Deno, 1989).

Uses of CBM

A plethora of uses for CBM exist in both special and general education settings. CBM can be used to screen for students who may need services, determine eligibility for services, program planning, progress monitoring, instructional decision-making, and program evaluation (Marston & Magnusson, 1985). More recently, CBM has been used to inform decisions about reintegration into mainstream classrooms (Powell-Smith & Stewart, 1998).

Progress Monitoring

Progress monitoring is one of the main uses for CBM (Deno, 1985). Much of the research in the area of progress monitoring has concentrated on the length of monitoring, difficulty of probes used, and expectations for student reading growth.

Reading time sample length. Deno, Mirkin, and Chiang (1982) compared differing time samples of reading for 45 students. Results showed a high correlation (.97) between mean numbers of words read correctly in 60 seconds and in 30 seconds. Fuchs, Tindal, and Deno (1984) also found a correlation of .97 between oral reading for 30 seconds and for 60 seconds. Although a high correlation was found, researchers indicated a slight advantage to oral reading fluency measured for 60 seconds rather than 30 seconds. Current practice uses a one-minute time sample as the standard for measuring student performance in reading.

Appropriate level of monitoring probes. Several studies have investigated the effects of passage difficulty on the sensitivity of CBM as a measure of student growth in reading (Hintze, Daly, & Shapiro, 1998; Marston & Magnusson, 1985; Shinn, Gleason, & Tindal, 1989). Shinn et al. varied the level of basal reading passages and measured student progress for 30 students in grades one through six who were reading below grade level. Participants were monitored four days a week for four weeks. Half of the participants were monitored in material one level above their instructional level and one level below their instructional level. The other half of the participants were monitored in materials that were two and four levels above their instructional level. For both groups easier passages resulted in significantly more words read correctly per minute, however no significant differences in slopes were found. These results indicate that the difficulty of the reading passage did not affect student progress. The researchers state that the failure to find differences in the slopes may have been due to the small sample size and low power.

Using a sample of 291 students in Minneapolis schools, Marston and Magnusson (1985) investigated the level of material that yielded the greatest slope. The researchers posited that the level of the material that yields the greatest slope is the most sensitive measure of student progress and thus the best choice for progress monitoring. Results showed the greatest slope for material in which the students read 11 to 20 words correct per minute for first and second graders and 41 to 50 words read correctly per minute for third through sixth graders.

In 1988, Marston and Magnusson adjusted these guidelines for the selection of appropriate level for monitoring progress. The reading level at which the student reads between 10 and 30 words correctly per minute was suggested for first and second grade students. Third through sixth graders should be reading between 30 and 60 words correctly per minute in monitoring material.

More recently, Hintze et al. (1998) investigated the effects of passage difficulty on student reading progress. Participants included 80 students in grades one through four. Participants were monitored twice a week for ten weeks at grade level and at the more difficult goal level. Results revealed higher slopes in grade level than goal level material for first and second grade students. There were no significant differences in slopes for third and fourth grade students. These results are consistent with the emphasis on decoding and fluency in grade level material in the lower grades, but not the higher grades. Students in first and second grade have exposure to the skills taught at grade level, but in higher grades reading skills of fluency are no longer emphasized or taught.

Consensus has not been reached in the literature regarding the difficulty level of reading probes for progress monitoring. Shinn et al. found no differences in reading growth due to differences in passage difficulty for students reading below grade level. However, Hintze et al. (1998) found a possible interaction between passage difficulty and student reading level. Despite the contradictory findings, CBM researchers traditionally recommend monitoring at a higher level than current instruction.

Expectations of reading growth. Several researchers have studied expected oral reading growth rates for different grade levels and types of educational placement (Fuchs, Fuchs, Hamlet, Walz, & Germann, 1993; Marston & Magnusson, 1985; Marston &

Tindal, 1995). Marston and Magnusson (1985) investigated weekly reading growth rates for students receiving services in regular education, Title I, and special education in grades one through three. The average weekly gains in words read correctly per minute decreased with increasing grade level and results are displayed in Table 1. Similarly, Marston and Tindal (1995) found that the average special education student gained 1.5 words per week in first and second grade. Results also showed that students in special education in grades three to six gained an average of only one word read correctly per week. In addition, students in regular education gained two to three words read correctly per week on average and students in grades three to six averaged 1.5 to 2.5 words gained per week.

Fuchs et al. (1993) used CBM ORF to monitor 117 students, 14 in special education classes, in grades one through six for one year. Participants were monitored once per week in grade level, non-curriculum material. Results indicated significant differences in the slope of reading growth as a function of grade. The mean gain in words read correctly per minute was the greatest for first graders and the least for sixth graders. Results are displayed in Table 1.

King, Bradley-Klug, Knoff, and Powell-Smith (2002) investigated expected growth rates of students in special education. The study participants were 56 students enrolled in specific learning disability (SLD) programs in three schools in Hernando County, FL. Probes for progress monitoring were developed from three different reading series. Participants were monitored two times per week for eight weeks in each of the three probe types. Participants were monitored in material one level above their instructional level. Results showed a significant main effect of curriculum on mean regression slopes.

The slope of improvement averaged across reading levels varied across different reading materials. There was also a main effect of instructional reading level on the oral reading rate mean regression slopes. However, unlike previous research (Fuchs et al., 1993; Marston & Tindal, 1995), King et al. did not find that students in lower grade instructional levels displayed higher oral reading growth rates. Instead, students at the first grade instructional level exhibited the lowest growth rates when reading rates across curricula were averaged. Students at the second grade reading level had the highest rate of progress.

Deno, Fuchs, Marston, and Shin (2001) investigated expected reading growth rates for 2675 general education students and 324 students receiving special education services in grades one through six. Data were derived from existing databases from four local education agencies from across the United States. The agencies varied in data collection with one agency collecting weekly data throughout the school year and the other three agencies collecting data only three times during the school year. All four agencies used grade level materials. Results indicated a general trend with decreasing mean weekly gains in words read correctly per minute with each successive higher grade level. The only exceptions to the trend were a slight increase in the mean words read correctly per minute from fifth to sixth grade for both general and special education and a stable mean increase in words read correctly per minute across third through fifth grade for special education. The special education mean gains were much lower than the general education gains for grades one trough four. However, nearly identical growth rates for general education and special education were found for grades five and six. Results are displayed in Table 1.

Table 1

Weekly Growth Rates in Number of Words Read Correctly Across Studies

Study	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Marston & Magnusson (1985)						
Regular Education	4.4	2.7	2.6			
Special Education	1.6	1.4	1.9			
Fuchs, Fuchs, Hamlet, Walz, & Germann (1993)						
Combined	2.1	1.5	1.1	.84	.49	.32
Deno, Fuchs, Marston, & Shin (2001)						
Regular Education	1.80	1.66	1.18	1.01	.58	.66
Special Education	.83	.57	.58	.58	.58	.62
King, Bradley-Klug, Knoff, and Powell-Smith (2002)						
Special Education	.0007[a]	.931[a]	.003[a]			

Note. [a]Data averaged across all three probe types used

Investigations of expected reading growth rates varied in results by grade level. Some of these differences may be due to the number of data collection points in each study. For example, Deno et al. (2001) data was collected only three times in a year, Marston and Magnusson (1985) collected data once a week for 16 weeks, and King et al. (2002) collected data twice a week for eight weeks. Differences also may be due to variability in participant populations. For instance, data for both special education and general education students was combined in the Fuchs et al. (1993) study, whereas Deno et al. (2001) and Marston and Magnusson (1985) separated the two groups to report results. In addition, the disparity in findings could be due to the difficulty level of the monitoring materials. King et al. (2002) monitored students at one level above their

instructional level, while Fuchs et al. (1993) and Marston and Tindal (1995) monitored in grade level material.

CBM Reading Reliability

Test-retest reliability. Several studies have established high test-retest reliability for CBM oral reading measures (Tindal, Germann, & Deno, 1983; Tindal, Marston, & Deno, 1983; Hartman & Fuller, 1997). Tindal, Germann, et al. randomly sampled 30 fifth grade regular education students from the six school districts in one county. Test-retest reliability was examined in two randomly selected passages from the basal reader in which each student was instructed. The basal curriculum differed by school. Each participant was tested twice with two weeks between sessions. The test-retest reliability of the CBM oral reading passages was .97.

Tindal, Marston, et al. (1983) investigated test-retest reliability using a much larger sample of 566 students in grades one through six. All participants were administered third grade passages from three different basal series. Participants were tested twice with 10 weeks between sessions. The test-retest reliability of the CBM passages was .92.

More recently, Hartman and Fuller (1997) examined the test-retest reliability of CBM oral reading using 26 students in grades one through three from a rural elementary school. Reading passages developed from literature used in the classrooms of the students were administered twice to each student with two weeks separating sessions. Results showed test-retest reliability for the three grade levels of students to range from .91 to .99.

Alternate forms reliability. Tindal, Germann, et al. (1983), Tindal, Marston, et al. (1983), and Hintze, Shapiro, and Lutz (1994) investigated alternate forms reliability. In

the study described above, Tindal, Germann, et al. also administered two parallel reading probes during the same session and found the alternate forms reliability to be .94. Additionally, Tindal, Marston, et al. administered alternate forms one week apart and found a reliability of .89.

Hintze et al. (1994) found somewhat lower alternate forms reliability. Participants were 48 third grade students from two rural schools. Half of the participants were instructed using a basal series and the other participants were instructed in a literature-based curriculum. Each participant was administered two different reading passages developed from each of the curricula. Alternate forms reliability ranged from .56 to .96. The lower reliability coefficients may be due to the different curricula used in measurement.

CBM Reading Validity

Criterion validity. To establish the criterion validity of the CBM one-minute oral reading measure, several studies have compared words read correctly to other measures of reading. These measures have included both other curriculum-based measures and published, standardized, norm-referenced measures.

Deno et al. (1982) conducted studies to determine concurrent validity using several measures. In one study, several curriculum-based and published, standardized measures were administered to 15 students with learning disabilities and 18 regular education students in grades one through five. The curriculum-based measurement materials included three word lists with 60 words on each (words in isolation), 600-word passages from three different basal series in which students read only every fifth word (words in context), 300-word oral reading passages from three basal series, and three

cloze comprehension passages from the same three basal series. The published and standardized measures included the Stanford Diagnostic Reading Test (SDRT) Reading Comprehension subtest (Karlsen, Madden, & Garner, 1975), and the Woodcock Reading Mastery Test (WRMT) Word Identification and Word Comprehension subtests (Woodcock, 1973). Results showed that the correlations between the one-minute oral reading passages and the other curriculum-based reading measures ranged from .81 to .89. The correlations between the oral reading passage and the published and standardized measures ranged from .73 to .87.

In another study, Deno et al. (1982) administered several measures to 23 students enrolled in a program for learning disabilities and 43 students enrolled in regular education. The measures used were lists of third and sixth grade words in isolation, an oral reading passage from a basal reader, a sixth grade level cloze passage, the Reading and Phonetic Analysis subtests of the Stanford Diagnostic Reading Test (Karlsen, Madden, & Garner, 1975), and the Peabody Individual Achievement Test (PIAT) Inferential and Literal Reading Comprehension subtests (Dunn & Markwardt, 1970). The correlations between the one-minute measure of oral reading and each of the other reading measures ranged from .71 to .87 with the exception of the word meaning measure with correlations ranging from .56 to .57.

Marston and Magnusson (1985) investigated the criterion validity of the CBM oral reading measure by comparing the number of words read correctly per minute to scores on published and standardized measures of reading for 26 third grade students. The measures used included the Vocabulary, Reading Words, Comprehension, Total Reading, and Word Study sections of the Stanford Achievement Test (Madden, Garner,

Rudman, Karlsen, & Marwin, 1973), the SRA Achievement series Vocabulary and Comprehension sections (Naslund, Thorpe, & Lefever, 1978), and reading probes developed from the Ginn 720 Reading Series (Clymer & Fenn, 1979). Correlations between the subtest scores and the number of words read correctly on the reading passage ranged from .80 to .90 with the exception of the vocabulary sections. The words read correctly were later compared to teacher ratings of student achievement on a scale from one to five. Teacher ratings and words read correctly were correlated at .77.

Hintze, Shapiro, Conte, and Basile (1997) studied the validity of the CBM reading techniques across different types of curricula. The 57 elementary students who participated in the study were instructed in a literature-based basal series with supplemental trade books. Oral reading probes were developed from the basal series and the authentic reading material. The participants read passages from each type of curriculum in each grade level one through five. The researchers compared the CBM oral reading results with scores from the Degrees of Reading Power Test (DRP). The DRP test (Koslin, Koslin, Zeno, & Ivens, 1989) is similar to a cloze reading comprehension measure with escalating difficulty of reading passages. The number of words read correctly from the reading probes and the DRP test scores correlated at .67 for the authentic literature probes and .64 for the basal series probes. These studies have provided a solid foundation of criterion validity for curriculum-based measurement oral reading fluency measures.

Construct validity. Fuchs, Fuchs, and Maxwell (1988) assessed an aspect of construct validity by comparing the average number of words read correctly per minute on CBM ORF with two subtests of the Stanford Achievement Test (1988). The Stanford

Achievement subtest scores and the CBM reading results for 10 boys aged nine to fifteen years with mild to moderate disabilities were compared. The researchers purported that a higher correlation between the CBM oral reading measure and the Stanford Achievement index of reading comprehension than between the oral reading and the index of study skills would indicate support for the claim that the CBM oral reading metric assesses reading comprehension skills. Fuchs et al. found that the correlation between the average number of words read correctly per minute on the CBM oral reading measure and the index of Reading Comprehension (.92) was significantly greater than the correlation between the CBM reading measure and the index of Word Study Skills (.81). These findings seem to lend support to the construct validity of the CBM oral reading fluency measure.

Discriminant validity. Several studies have investigated discriminant validity by assessing the degree to which the oral reading CBM measure can discriminate among groups that theoretically differ in their reading skills. Deno et al. (1982) provided evidence for CBM oral reading discriminant validity by comparing the performance of 27 students in regular education to 18 students in a learning disabilities resource program in grades one through six. The mean number of words read correctly was higher for general education students than for students in the learning disabilities resource program. The differences between the group means remained fairly constant across grade levels.

A similar study by Deno et al. (1982) compared the oral reading rates of 43 general education students and 23 students enrolled in a learning disabilities resource program in first through sixth grades. Oral reading scores for the general education

students were three times higher than for the students in the learning disabilities resource program.

Shinn and Marston (1985) studied discriminant validity by comparing the oral reading performance of 51 general education students, 110 students receiving Chapter 1 services, and 48 students with mild disabilities receiving part-time special education services. The participants were all in the fourth to the sixth grades. Analyses of variance revealed significant differences among all the groups in the mean number of words read correctly. The general education students attained the highest mean of words read correctly followed by the mean for students receiving Chapter 1 services. The mean number of words read correctly by students receiving special education services was significantly lower than the means for the other 2 student groups.

Marston and Magnusson (1985) provided evidence for CBM oral reading discriminant validity, but some of the results were mixed. They compared the oral reading performance of 130 general education students, 24 special education students, and 104 students in a Title 1 program who were all in the first through third grades. Progress monitoring lasted for 16 weeks with passages from each student's grade level. The group mean of the general education students was significantly higher than the means for Title 1 or special education students. Although, there was no significant difference between the group means for the special education and Title 1 students, when analysis was done by grade level, significant differences were found between the groups at grades two and three. Differences in the developmental trends of the three student groups were assessed by mean slope comparisons. Discrimination between the three groups was shown in grades one and two with general education students averaging the

most words gained per week and special education students averaging the fewest words gained. For grade three, the slope for general education students was again the highest, but the slopes for special education and Title 1 students were the same. Overall, oral reading measurement was able to reliably discriminate between students who were in special education programs, Title 1 or Chapter 1 programs, and students who were not in such programs.

Another way that researchers have investigated discriminant validity is through the assessment of the CBM oral reading measure's sensitivity to growth. In this type of research, developmental growth rates are assessed as students at various grade levels read passages of equal difficulty. The number of words read correctly should vary positively according to student grade level. This would allow student progress in reading to be closely monitored and utilized by school staff.

Deno et al. (1982) studied developmental growth patterns for oral reading across grades one through six for students in regular education and in a learning disabilities resource program. Passages developed from both third and sixth grade material were utilized. The researchers found that the slope of oral reading fluency was fairly linear in a positive direction across grade levels for both student groups regardless of the passage difficulty.

Fuchs and Deno (1992) examined oral reading growth rates to determine if they were affected by passage difficulty or the curriculum from which they were drawn. The 91 study participants in grades one through six all read passages developed from the third grade material of two different reading series. The developmental slopes for both curriculum series were parallel, stable, and linear. The only exception was the

performance drop for fifth grade students. To assess the effect of passage difficulty on developmental growth rates, Fuchs and Deno (1992) calculated slopes for material developed from each grade level in each curriculum series. Results showed a slight decrease in developmental growth rates as the difficulty of the material increased. These findings suggest that CBM oral reading developmental growth rates are not affected by type of curriculum and only somewhat affected by the difficulty of the material. Table 2 contains a summary of the cited discriminant validity research. Overall, CBM ORF was able to reliably discriminate between students who were in special education programs, Title 1 or Chapter 1 programs, and students who were not in such programs. In addition, CBM ORF was able to validly discriminate growth in reading proficiency.

Face validity. Yell, Deno, and Marston (1992) surveyed special education teachers to investigate the barriers to the use of CBM ORF. One primary concern of the teachers was whether this metric measures comprehension. Supplementing oral reading with a measure of comprehension was suggested. Although other empirically sound curriculum-based measures of reading have been shown to measure comprehension (i.e., cloze and written recall), none of them correlate as highly with standardized measures of reading comprehension as the oral reading fluency measure (Shapiro, 1989). According to Fuchs and Deno (1992, p. 233), "Reading aloud from a text demonstrates the strongest relation with socially important, widely used criterion measures of reading." The ability of the oral reading fluency metric to measure comprehension has been well documented (Deno, Mirkin, & Chiang, 1982, Marston & Magnusson, 1985) and results lend themselves well to instructional planning and databased decision-making.

Reading Probes from Varying Curriculum

The methods of curriculum-based assessment share three main tenants: (1) material is developed from the school or class curriculum, (2) measurement is repeated over time, and (3) informs instructional decision-making (Tucker, 1987). Over the past decade research has begun to question Tucker's first tenant. Fuchs and Deno (1994) concluded that it is not critical for instructionally useful measurement material to be developed from the students' curriculum. Instead the essential features of instructionally useful material are: (1) the opportunity for measurement to be repeated over time with maintenance of difficulty levels, (2) the ability to assess changes in actual outcomes of student achievement, and (3) the ability to produce quantitative as well as qualitative descriptions of student performance.

According to Fuchs and Deno (1994), the sole advantage of using probes based in the curriculum is face validity for teachers using CBM. Teachers most often develop tests from their own curriculum to assess student progress (Fuchs & Deno, 1994). Several disadvantages of using probes based in the curriculum were noted including the fact that curriculum materials within a program or book vary significantly with regard to difficulty. According to Fuchs, Fuchs, and Deno (1982), up to 10 reading passages may need to be drawn from a text before one passage may represent the text accurately. This type of variability in curricular materials introduces additional measurement error. Materials drawn from outside the curriculum may be more easily controlled for difficulty level. Secondly, when material is drawn from the curriculum in which the student is instructed, familiarity with passages may be a factor in the results. Student performance would be a result of prior experience with the material and not the application of learning.

Utilization of material with varying familiarity for the student can be a source of measurement error. Also, using materials based in the curriculum might reduce the generalizability of the information gained. Fuchs and Deno (1994) noted that if growth is demonstrated solely in curriculum material, decisions resulting from this information might not be generalizable to reading outside the curriculum. Vocabulary and linguistic patterns are recurrent in many curricula and may have a controlling effect on student performance. Students may perform well within the curriculum, but struggle with reading from other sources (Fuchs & Deno, 1994).

Few studies have been conducted to investigate the impact of curricular variation on curriculum-based measurement oral reading fluency. Fuchs and Deno (1992) conducted a study to assess the effects of the use of different curriculum materials in CBM and to add to the literature on CBM's technical adequacy. The participants in the study included 91 students in first through sixth grade. Sixteen percent of the participants received special education services and 25% received Chapter 1 services. A published and standardized achievement test (Passage Comprehension test of the Woodcock Reading Mastery Tests) and the CBM oral reading measure with probes developed from the Ginn 720 (1976) and probes developed from Scott-Foresman Unlimited reading series (1976) at each grade level from primer to sixth grade were administered. CBM reading showed durability of technical adequacy across different curriculum materials. Despite differences between the reading series (Scott-Foresman, a literature-based series with a comprehension focus and the Ginn Series with a more traditional diverse phonetic focus), no significant differences between the two series on the CBM were found. Slopes

computed for the two series curricula were comparably linear and stable. In addition, criterion validity was not affected by the difficulty of the passages used.

Hintze, Shapiro, and Lutz (1994) conducted a study to build upon the Fuchs and Deno (1992) study and to increase the literature base on the technical features of the CBM in reading. Participants were 48 third grade students, 24 students from a rural school and 24 students from an urban school in New Jersey. The students in the rural school were instructed with the Scott-Foresman reading series (literature-based) and urban students were instructed with the Houghton Mifflin basal series (traditional, phonics-based series). Students receiving special education or Chapter 1 services were not included in the sample. Probes were developed only from third grade material in each of the reading series. Researchers conducted teacher interviews and administered probes twice per week for nine weeks. Regardless of the reading series in which the students were taught, performance declined when students were monitored using the literature-based series and performance increased when monitored with the traditional basal series. Therefore, the use of probes developed from the traditional basal series for progress monitoring showed greater sensitivity to growth over time than the use of literature-based probes. In addition, students instructed using the traditional series showed significantly more improvement over time than students instructed using the literature-based series. The results of this study may have been limited by the non-random assignment of conditions and possible lack of instructional integrity as evidenced by the teacher interviews. Results of this study indicate that the curriculum material from which the reading passages are drawn may influence CBM oral reading results. These limitations may threaten the validity of the research. Additionally, the use of only third

grade students as participants may limit the generalizability of the findings to students at other grade levels.

In follow-up research, Hintze and Shapiro (1997) sought to extend Hintze et al.'s (1994) findings by including participants at various grade levels from a variety of schools. Participants of the study were 160 students in second through fifth grades from one urban and one suburban school district. Students in the suburban school district were instructed with the Scott-Foresman literature-based reading series and students in the urban district were instructed using the Houghton Mifflin basal series. Researchers conducted teacher interviews and developed CBM reading probes from each of the two instructional reading series. Student progress was monitored in each probe type two times per week for eight weeks. Students were monitored with probes developed from "challenging level materials" in which the student was proficient one grade level above the student's current grade level. When monitored using the literature-based probes, students displayed growth over time despite differences in materials used for instruction. Also, students in all grades except grade 2 displayed growth over time on the basal series probes. Larger growth rates were found for second, third, and fourth grade students when they were monitored with literature-based probes. In contrast, fifth grade students showed higher growth rates when monitored with the basal series probes. Despite these differences, both types of probes showed sensitivity to growth in oral reading rate over time irrespective of the materials in which students were instructed. In contrast to the Hintze et al. (1994) findings, positive slopes were found for all students when monitored in literature-based materials regardless of instructional curriculum.

Results of a study by Bradley-Klug, Shapiro, Lutz, and DuPaul (1998) supported the finding that literature-based probes may underestimate student performance (Hintze et al., 1994), but found that literature-based and basal-based probes were both sensitive to growth over time. Twenty-eight second graders and 30 fifth graders from three schools in suburban Pennsylvania participated in the study. Participants were randomly selected from students scoring between the 25th and 75th percentile on the California Test of Basic Skills. CBM reading probes were developed from the literature-based reading series in which the students were instructed (Houghton Mifflin Literary Series) and from a basal series (Harcourt Brace Jovanich Reading Series). Both types of probes were selected from end of the year goal materials as indicated by the second and fifth grade teachers. Progress was monitored two times per week for ten weeks. Overall, researchers found the CBM oral reading fluency measure to be sensitive to growth across time for students being instructed in a literature-based curriculum. Literature-based probes and basal series probes were both effective measures of student progress over time. Although both probe types displayed sensitivity, mean performance using basal probes was significantly higher than mean performance using literature based probes. The literature-based probes may have underestimated the level of student performance. In addition, the literature-based probes may have contained a greater percentage of unfamiliar vocabulary than the basal probes. Generalizability of these findings may be limited because only second grade students who were reading in the average range were included in the sample.

Powell-Smith and Bradley-Klug (2001) conducted research to expand prior research on the use of different sources used for the development of CBM reading probes and the use of these probes to monitor oral reading fluency over time. Participants were

36 students in the second grade whose reading levels were below those of their peers, according to their teachers. Students receiving special education services at the time of selection were not included in the sample. Researchers used reading probes developed from the basal series used in the students' school (Macmillan or Scribner) and TORF (Tests of Reading Fluency) developed by Deno, Deno, Marston, and Marston (1987). The TORF probes were not constructed from a specific curriculum. Probes used for progress monitoring were developed from long term goal material defined as one grade level above the student's current reading instructional level determined by SLA. Data collection involved administration and scoring of each type of probe twice per week for five weeks.

Progress monitoring with the two types of probes revealed a level change, but the rank order of individuals remained constant. The participants read significantly more words correct on the TORF probes than on the basal probes. However, there were no significant differences between the slopes of the scores for the two probe types. Both types of probes showed consistent sensitivity to growth in oral reading rates over time.

King, Bradley-Klug, Knoff, and Powell-Smith (2002) investigated the effects of progress monitoring materials developed from varying curricula when used with students receiving special education services. The study participants were 56 students enrolled in specific learning disability (SLD) programs in three schools in Hernando County, FL. Participants received primary reading instruction from an ESE teacher using the SRA Reading Mastery Series. Participants were separated into three groups by reading instructional level using survey level assessment. Group one read on a first grade instructional level. Groups two and three read on second and third grade levels,

respectively. Probes for progress monitoring were developed from the SRA Reading Mastery Rainbow Edition (Engelmann & Bruner, 1995), Scott Foresman Celebrate Reading! (1995-1997), and the Accelerated Reader Program (Advantage Learning Systems, Inc., 1996). The SRA Reading Mastery program is a direct instruction curriculum with controlled vocabulary presentation and sequence, while Celebrate Reading! is a literature-based series that is less controlled and sequenced. The Accelerated Reader Program is an authentic literature program with hundreds of books at predetermined levels of readability and no required sequencing. Participants were monitored two times per week for eight weeks in each of the three probe types. Participants were monitored in material one level above their instructional level.

Results of the King et al. (2002) research showed a significant main effect of curriculum on mean regression slopes. The slope of improvement averaged across reading levels was highest for the Accelerated Reader Program authentic literature probes and lowest for the Scott Foresman literature-based probes. There was also a main effect of instructional reading level on the oral reading rate mean regression slopes. However, unlike previous research (Fuchs et al., 1993; Marston & Tindal, 1995), King and her colleagues did not find that students in lower grade instructional levels displayed higher oral reading growth rates. Instead, students at the first grade instructional level exhibited the lowest growth rates when reading rates across curricula were averaged. Students at the second grade reading level had the highest rate of progress. One explanation for this disparity in findings is that the students were grouped by instructional level in the King et al. study and by grade level in the Fuchs et al. (1993) and Marston and Tindal (1995) studies. King et al. also found that the mean regression slopes differed significantly as a

function of instructional level when probes from either the direct instruction or the literature-based curricula were used. However, probes developed from the authentic literature curriculum worked equally as well for students at all three instructional levels.

Although differences between the probe types were found for the mean number of words read correctly, oral reading probes developed from varying curricula have shown sensitivity to reading growth over time (Hintze & Shapiro, 1997; Bradley-Klug et al., 1998; Powell-Smith & Bradley-Klug, 2001). However, King et al. (2002) found varying sensitivity to reading growth when different measurement materials were used for progress monitoring. In contradiction to other findings Hintze et al. (1994) reported that reading materials from outside the students' curriculum were not sensitive to CBM ORF growth over time. Studies investigating progress monitoring materials from different curricula have varied according to the length of progress monitoring data collection and the difficulty level of the materials used for monitoring. Additionally, research investigating the appropriate use of varying curricula for ORF progress monitoring has utilized curricular materials developed from only a few reading series and non-curricular materials. It is unclear whether findings from this research can be generalized to progress monitoring with materials developed from other sources.

Purpose of the Current Study

The purpose of the current study is to clarify and extend previous research on the comparibility of CBM ORF results using reading materials from outside of the students' curriculum for repeated measurement over time. Specifically, this study evaluated the use of generic measurement materials for progress monitoring student reading growth.

Additionally, this study investigated expected gains in words read correctly per minute over time at different grade levels.

Research questions addressed in this study were as follows:

1) Do curriculum-based probes result in the same number of words read correctly per minute as standardized probes developed outside of a curriculum?

2) Do curriculum-based probes result in the same rate of increase over time in words read correctly per minute as standardized probed developed outside of a curriculum?

3) Does the rate of increase over time in words read correctly per minute differ at different grade levels?

4) Is there an interaction effect of type of probe and grade level for the rate of increase in words read correctly per minute over time?

This study offers three main contributions to the present CBM ORF literature base. First, this study investigated the comparibility of two sources of measurement material (i.e., AIMSweb, Open Court) for CBM ORF progress monitoring that have not been previously examined. Secondly, by collecting progress monitoring data for a longer period of time (i.e., 19 weeks) than other studies examining the relationship between reading growth over time and the material used to measure it (i.e., five to ten weeks), this study displayed a more complete picture of reading growth over time. The results of this research helped to clear up the disparity in findings. Third, by examining differences in expected growth rates due to differences in measurement materials, this study added to the research on academic growth standards for students and clarified reasons for differences in the results of previous research. In addition, this research may have a

practical impact in that it may inform decisions about the choice of progress monitoring materials and assist in the development of appropriate goals for improvement in ORF with the establishment of academic growth standards in varying curricula.

Chapter Three

Method

This study addressed the effect of the source of reading material for progress monitoring (reading probe type) on student reading performance over time within CBM. Specifically, the number of words read correctly per minute (WRC) using reading probes sampled from the student's curriculum was compared to WRC using generic probes. The rate of growth over time in WRC with each type of probe at different grade levels was also compared.

Setting

Data for this study were collected from January to May of 2002. Data collection was conducted at a central Florida charter school serving approximately 80 children who have had frequent out-of-school suspensions and/ or expulsions from central Florida public schools. Each teacher collected data in her individual classroom. The reading curriculum used for instruction was the SRA Open Court basal series (1997). The Open Court basal series is grounded in systematic phonemic awareness, word knowledge, and comprehension skills along with writing and language arts skills and strategies. Curriculum-based measurement with probes developed from Open Court material has been used at this school for two years.

Overall Sample and Missing Data

Ninety-three students attended the charter school at the time of data collection. However, not all student data were used in analysis. Sixteen students did not read at least 10 words per minute in grade level materials and thus progress monitoring was completed with an alternate assessment tool. Therefore, data from these 16 participants were not used in analysis. Secondly, the data from two students were not used in analysis due to a change in the level of monitoring materials during data collection (e.g., began with grade 2 materials and finished with grade 3 materials). Lastly, the number of data points obtained for each participant was calculated. When greater than 50% of the data points were missing for any participant, those participants' data were dropped from subsequent analyses. According to data collectors, greater than 50% of missing data resulted from the student changing schools. This criterion resulted in 11 participants' data being dropped from the study.

A total of 29 participants' data were dropped from analyses. These participants were eight girls and 21 boys, 24 African-American students, three White students, and two Hispanic students. Fifteen of the dropped participants were on free or reduced lunch status. Twenty of the students were in first grade, six in second grade, and three in third grade.

Description of the Final Sample

The final sample for data analyses consisted of 64 students, 41 (64.1%) boys and 23 (35.9%) girls, 26 (40.6%) first graders, 21 (32.8%) second graders, and 17 (26.6%) third graders. Sixty-one (95.3%) participants were African-American and 3 (4.7%)

participants were white. Table 2 shows the demographics within the grade level groups.

Table 2

Participant Demographics by Grade Level

	Grade 1		Grade 2		Grade 3	
	N	%	N	%	N	%
Gender						
Male	16	61.5	13	61.9	12	70.6
Female	10	38.5	8	38.1	5	29.4
Ethnicity						
African-American	25	96.2	20	95.2	16	94.1
White	1	3.8	1	4.8	1	5.9

Materials

Open Court reading probes. The Open Court basal series (1997) is a highly structured and explicit phonics-based approach to teaching reading that organizes reading passages according to reading difficulty and phonetical concepts. CBM reading probes were developed at each grade level by randomly selecting from material at the beginning, middle, and end of the basal readers for each grade level.

AimsWeb reading probes. AimsWeb probes are generic measurement materials that were written by educators in one school district in the Midwest. Howe and Shinn (2001) conducted statistical analyses on a pool of possible reading passages using data from a sample of 24 children per grade level from a mid-western school district. Passages with alternate-form reliability below .70 when compared with other passages at the same grade level were discarded. Additionally, passages with a mean WRC that differed from the grade-level mean by more than one standard deviation were also discarded. Lexile-graded standards of readability (Stenner, Smith, & Burdick, 1983)

were used to estimate reading passage difficulty. Passages with readability scores outside the normal range for each grade level were eliminated. Finally, 23 to 33 probes in each grade level qualified to be used for progress monitoring. The average alternate form reliability of the AimsWeb progress monitoring probes for grades one through three is .86 (Howe & Shinn, 2001).

Progress monitoring was conducted with students in their current grade level material using both Open Court and AIMSweb probes. Open Court probes were selected at random for use in progress monitoring, while AimsWeb probes were used in the order recommended by the developers.

Dependent Measures

The number of words read correctly was determined for each participant in each probe type two times per week for 19 weeks. The mean number of words read correctly for the two weekly administrations in each probe type were used in analysis so that each participant had one data point per week in each probe type for the 19 weeks.

The number of words read correctly per minute was determined using the following scoring criteria for CBM in reading described by Shinn (1989):

1. *Words read correctly*. Words read correctly are those words that are pronounced correctly, given the reading context.
    a. The word "read" must be pronounced "reed" when presented in the context of "He will read the book," not as "red."
    b. Repetitions are not counted as incorrects.
    c. Self-corrections within 3 seconds are counted as correctly read words.

2. *Words read incorrectly*. The following types of errors are counted: (a) mispronunciations, (b) substitutions, and (c) omissions. Further, words not read within three seconds are counted as errors.
    a. *Mispronunciations* are words that are misread: *dog* for *dig*.
    b. *Substitutions* are words that are substituted for the stimulus word; this is often inferred by a one-to-one correspondence between word orders: *dog* for *cat*.
    c. *Omissions* or words skipped or not read; if a student skips an entire line, each word is counted as an error.
3. *3-second rule*. If a student is struggling to pronounce a word or hesitates for three seconds, the student is told the word, and it is counted as an error. (p. 239-240)

The dependent variables in this study were level, which was defined as the mean of the data points for each type of probe, and the slopes derived from Ordinary Least Squares Linear Regression Equations for the number of words read correctly across all of the data collection days. The slope of the number of words read correctly for each student in each probe type was calculated using an Ordinary Least Squares regression. Each slope reflected the student's gain in the number of words read correctly in each type of probe over time. The slopes were calculated based on the number of probes given to each student, which differed due to absences on the day of assessment or a change in schools. Students who had fewer than 20 data points were not used in analysis. The variables used in this study are listed in Table 3.

Table 3

Types of Variables Used in Study

Variable	Variable Type	Scale of Measurement
Materials used to monitor student reading progress (Type of probe, AimsWeb or Open Court)	Categorical Independent Variable	Nominal
Level of oral reading performance (mean of all data points)	Continuous Dependent Variable	Interval
Slope of oral reading performance (determined by Ordinary Least Squares regression)	Continuous Dependent Variable	Interval
Grade level (one, two, or three)	Categorical Independent Variable	Ordinal

Procedure

Data collectors. Each of the teachers at the school collected data for each student in their classes. Teachers were each female and African-American. Teachers chose the time of day and location in the classroom. Therefore these variables were different in different classrooms.

Teachers were trained in the use of CBM by University of South Florida faculty during a half-day session. Additionally, a power-point presentation provided by Edformation, Inc. was used to train teachers in the use of AIMSweb probes and the input of information into the computer. Interrater agreement was collected prior to the beginning of data collection as part of the Edformation training for the use of AimsWeb materials and met Edformation training standards.

Progress monitoring. Progress monitoring was conducted twice a week in each type of probe for each of the 19 weeks of the semester from January to May of 2002. Teachers each collected data in their own classrooms. Two Open Court probes and two AimsWeb probe were used each week and the administration order was Open Court and then AIMSweb each time. The mean of the two weekly administrations for each probe type was used in analysis. The purpose of progress monitoring was to determine each student's rate of progress and the differences in student performance in the two probe types.

Each student read the passage out loud according to standardized directions. Each teacher asked an individual student to read a passage and timed the student for one minute with a stopwatch. Teachers used a copy of the student probes to mark each student's errors and asked the students to stop reading after one minute. Teachers counted the number of words read correctly and the number of errors and recorded these numbers on the teacher's copy of each probe.

Research design. A repeated measures design was used in this study. The independent variables were the type of probes used in progress monitoring and the students' grade level. The dependent variables were the level of the number of words read correctly and the slope (rate of change in WRCPM per week) in each type of probe.

Data analysis. The data analyses was used to answer the following questions:

1) Do curriculum-based probes result in the same number of words read correctly per minute as standardized probes developed outside of a curriculum?

2) Do curriculum-based probes result in the same rate of increase over time in words read correctly per minute as standardized probed developed outside of a curriculum?

3) Does the rate of increase over time in words read correctly per minute differ at different grade levels?

4) Is there an interaction effect of type of probe and grade level for the rate of increase in words read correctly per minute over time?

The first step in analysis was to plot the WRCPM data for each participant in each type of probe over the twenty weeks. Each mean weekly data collection point was plotted with WRC on the y-axis and date on the x-axis. This study conceptualized CBM ORF growth as a linear function using the slope as the parameter describing the change. The assumption of linearity was based on the results of several studies that confirm reading growth over time as a linear relationship (e.g., Deno, Fuchs, Marston, & Shin, 2001; Shinn, Good, & Stein 1989). A regression line was fit to each student's data using Ordinary Least Squares Regression and the r^2 statistic was used to judge the goodness of fit and the appropriateness of a linear relationship. Ordinary Least Squares regression was used because it has been shown to estimate slope more precisely than the split-middle method and it minimizes the distance between the data points and the regression line (Shinn et al., 1989). The level and slope of the regression line for each participant in each type of probe were used in analyses. The level was determined by taking the mean of all of the data points for each individual for each type of probe and represents the average number of words read correctly. The slope represents the rate of change in the

number of words read correctly over time. The average level and slope for each grade level and for all participants combined were determined.

The intercept for each regression line was calculated using Ordinary Least Squares regression to assess students' initial reading performance for each type of probe. The correlation between the slopes and intercepts for each grade level and an overall correlation was examined to determine if differences in slope values were related to differences in students' initial level of performance.

Two repeated measures ANOVAs were used in this study because participants experienced each of the probe type conditions. A 3 (grade) x 2 (probe type) repeated measures ANOVA using the three grade levels as a between group variable and the two probe types as a within group variable was conducted for the slopes as well as for the levels. This procedure was used to determine if there was an effect by grade and/or probe type and also if there was an interaction between grade and probe type. Post hoc analyses were conducted to further analyze ANOVA results.

Chapter Four

Results

This chapter describes the participant sample, the statistical tests used to answer the research questions, and the outcomes of the various analyses. Each research question will be directly addressed.

Individual Analysis

Data were plotted for each individual in each type of probe over the 19 weeks with time in weeks on the x-axis and mean number of words read correctly per minute (WRCPM) on the y-axis. Data for 17 of the 19 weeks were plotted and used in the analyses because no data were collected in week 11 due to spring vacation. Also, preparation for state standardized testing was conducted during week seven and data were collected for only 12.5% of participants during that week. Thus, data for week 7 were not used in analyses either.

A regression line was fit to each student's data using Ordinary Least Squares regression. The R^2 statistic was used to judge the goodness of fit and the appropriateness of a linear relationship for each individual. Descriptive data for each probe type in each grade are reported in Table 5. These results showed that the goodness of fit for the individual regression lines was nearly the same for the two types of probes overall and in first grade. However, the goodness of fit was slightly better for second grade students

using the Open Court measurement materials than the AimsWeb materials. Use of either type of material for third grade student progress monitoring resulted in overall low goodness of fit for individuals in the study.

Table 5

R^2 Descriptives by Grade Level and Type of Probe

	AimsWeb		Open Court	
Grade	Mean	SD	Mean	SD
1	0.62	0.24	0.61	0.23
2	0.37	0.26	0.44	0.23
3	0.13	0.12	0.06	0.08
combined	0.41	0.30	0.41	0.30

All participant data were plotted for each probe type and are displayed in Figures 1 and 2. To further examine the differences due to measurement material, data were plotted for each grade level in each type of probe and are displayed in Figures 3, 4, 5, 6, 7, and 8. Although the aggregated data seem similar, increased difference between the probe types is more apparent for third grade. The number of outliers increased in third grade, the standard deviation increased, and the slopes leveled out. Further analyses were necessary to determine if using different probe types for measurement made a significant difference in the resulting WRCPM.

Level Calculation

The level was calculated by taking the mean of the words read correctly per minute (WRCPM) for each participant in each type of probe. The levels were also calculated across grades. The means and standard deviations for each probe type at each grade level are reported in Table 6.

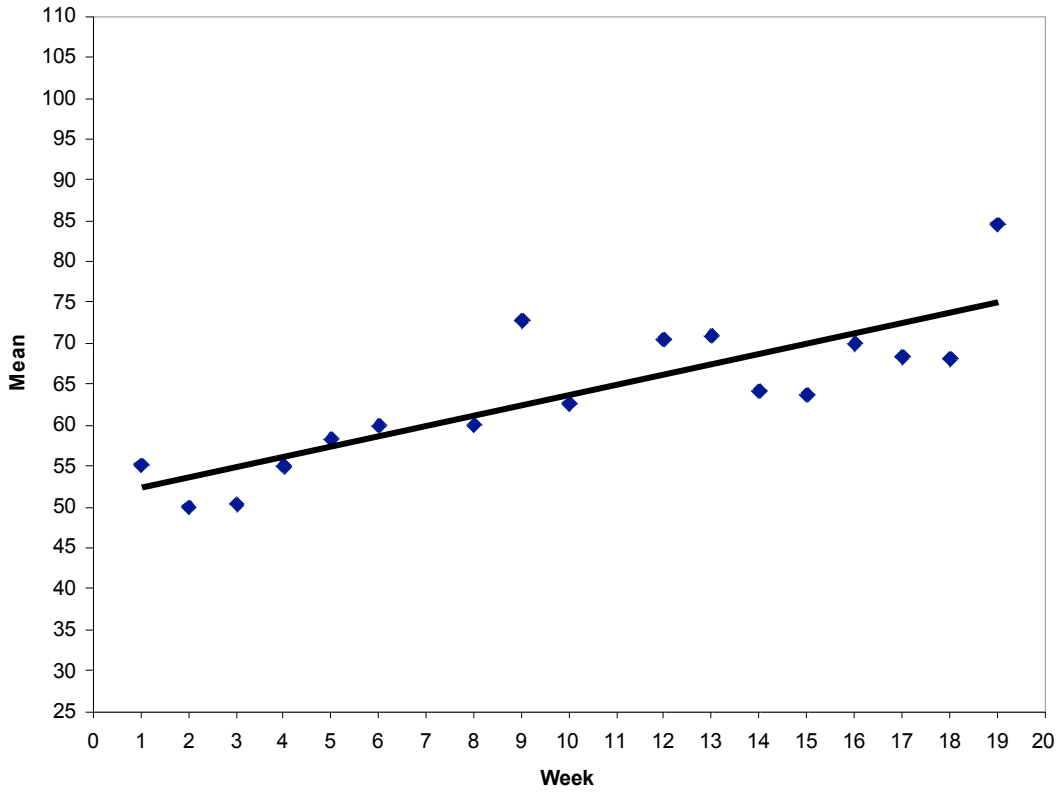

Figure 1. Overall mean WRCPM measured with AimsWeb materials

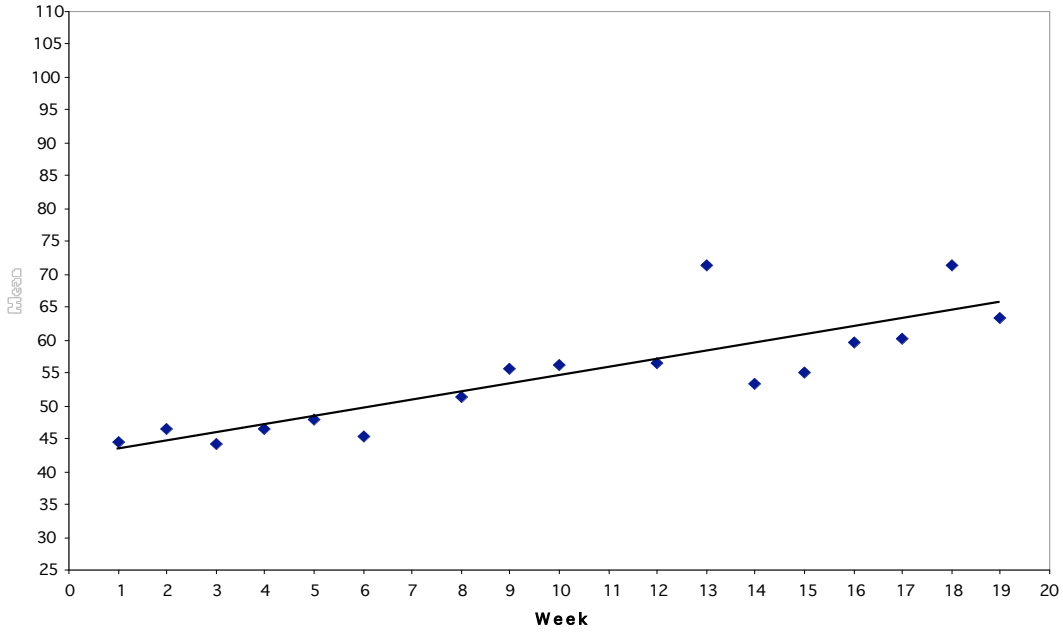

Figure 2. Overall mean WRCPM measured with Open Court materials

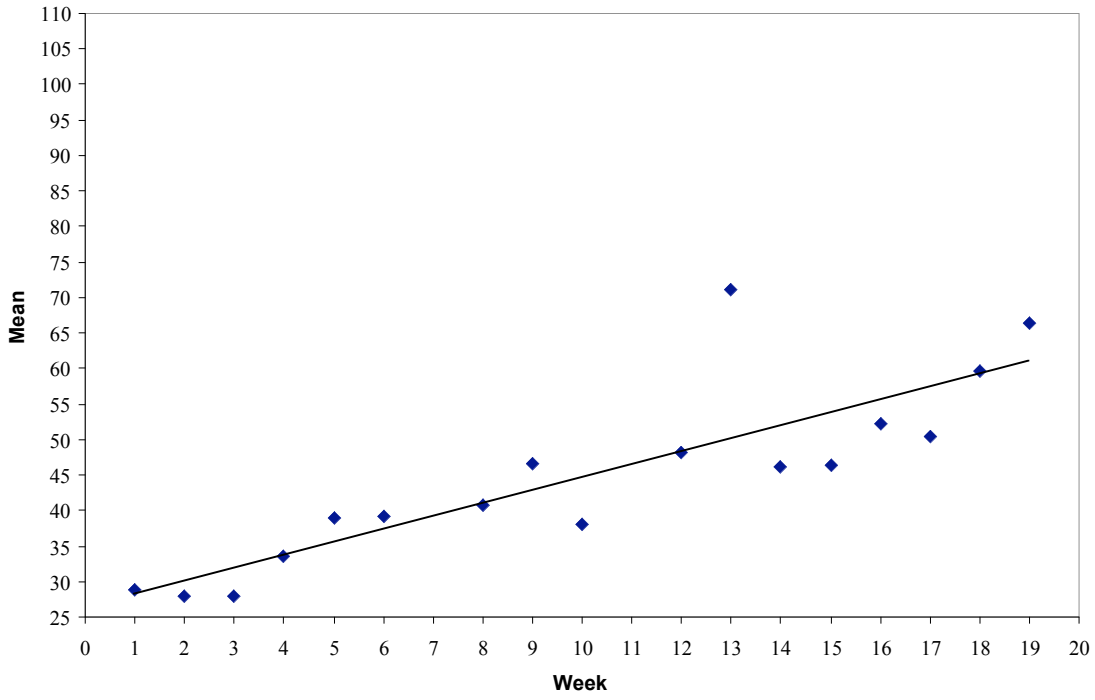

Figure 3. First grade mean WRCPM measured with AimsWeb materials

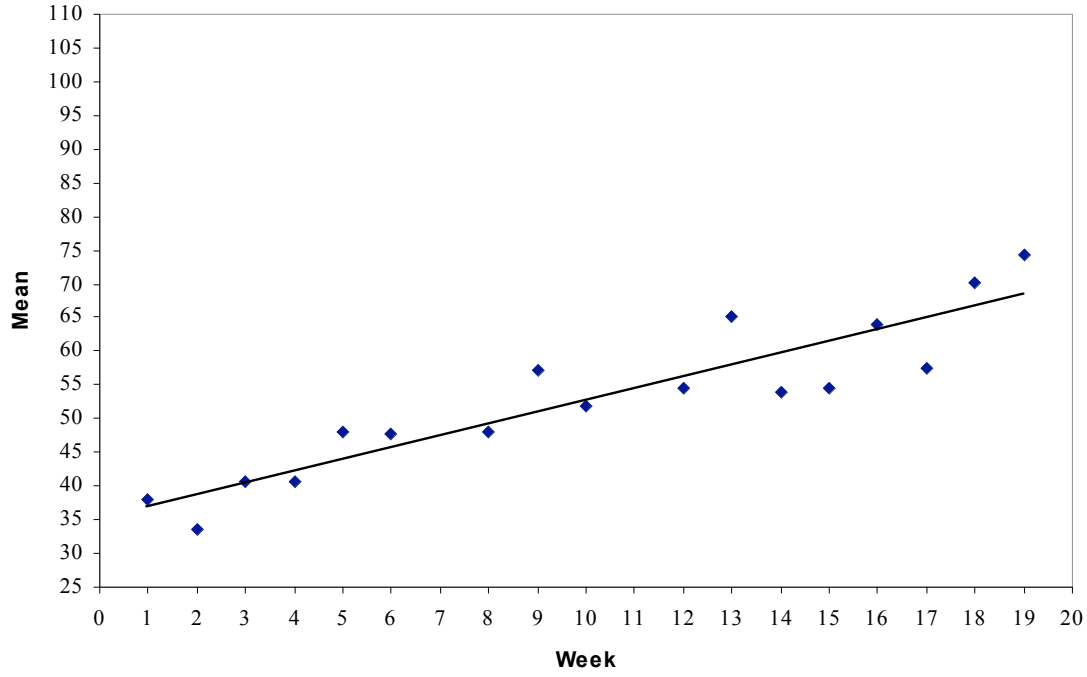

Figure 4. First grade mean WRCPM measured with Open Court materials

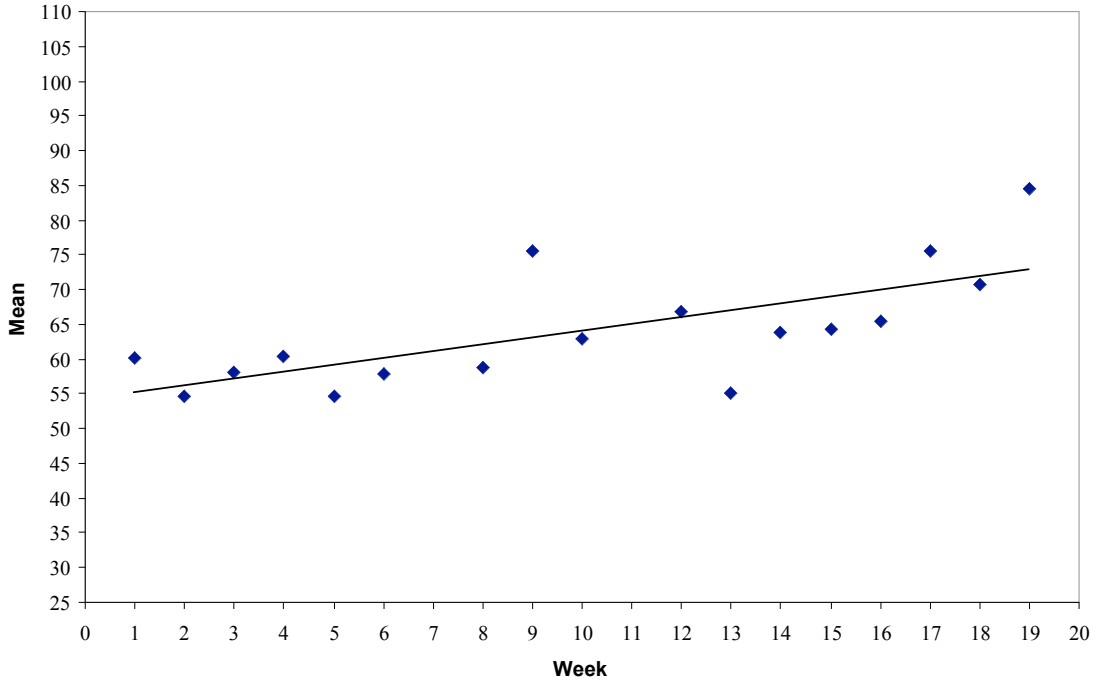

Figure 5. Second grade mean WRCPM measured with AimsWeb materials

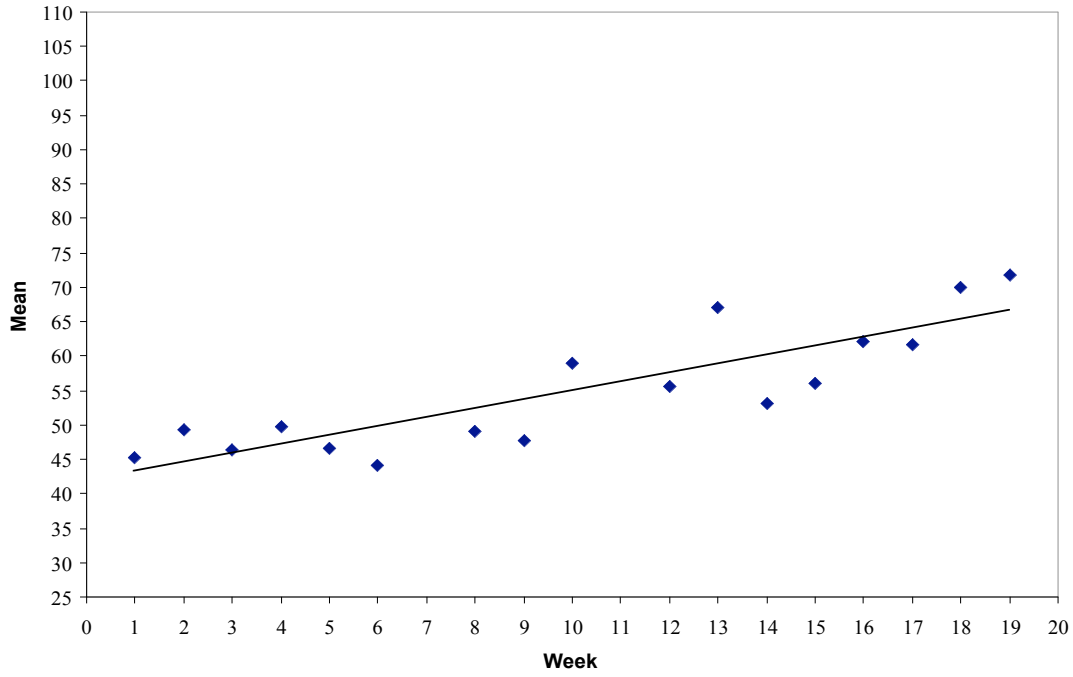

Figure 6. Second grade mean WRCPM measured with Open Court materials

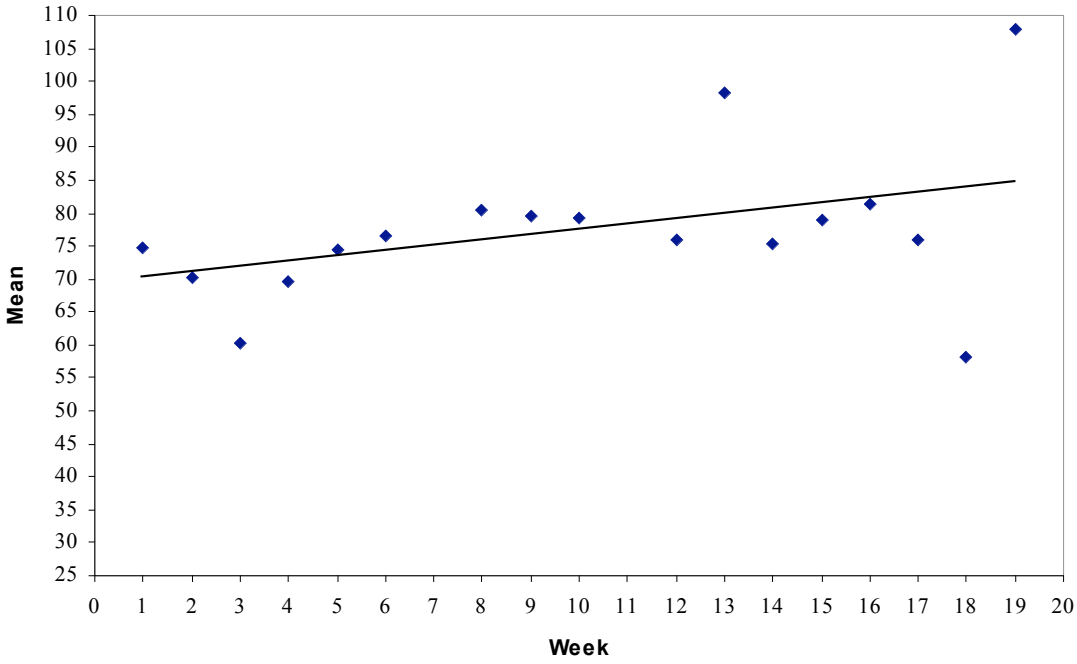

Figure 7. Third grade mean WRCPM measured with AimsWeb materials

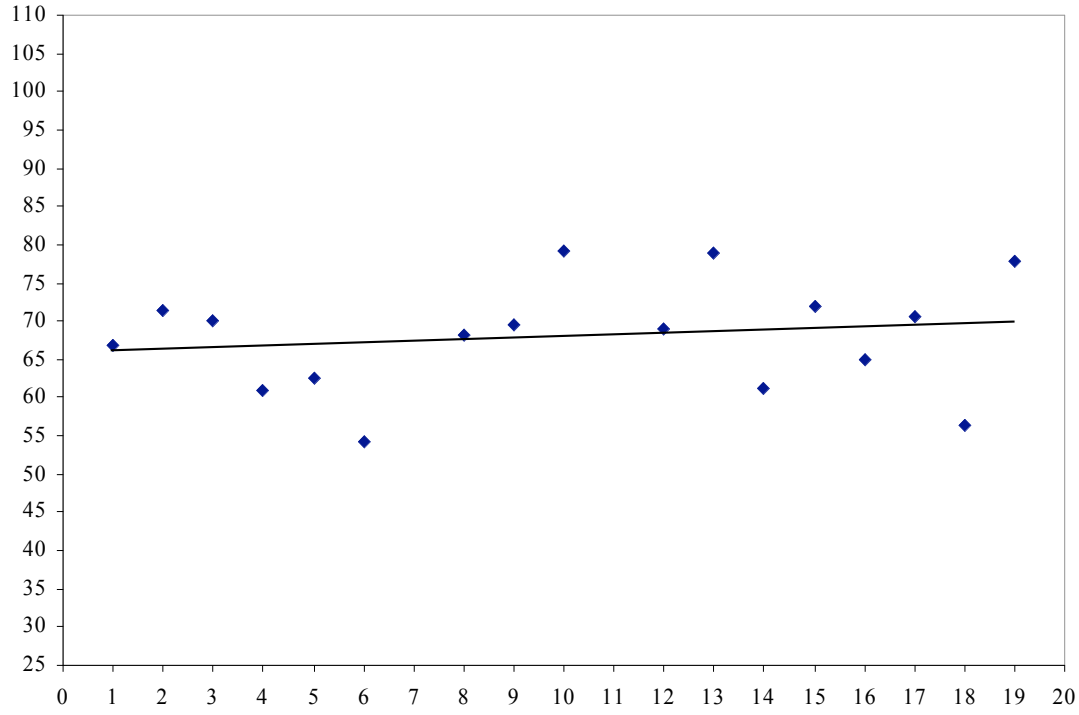

Figure 8. Third grade mean WRCPM measured with Open Court materials

Table 6

Level Descriptive Statistics by Type of Probe and Grade

		Open Court probes		AimsWeb probes	
Grade Level	N	Mean	SD	Mean	SD
1	26	27.18	14.46	35.35	19.25
2	21	41.07	22.17	54.49	30.57
3	17	66.22	15.66	71.84	21.21

Pearson Correlations Between Open Court WRCPM and AIMSweb WRCPM by week

In addition to descriptive analyses Pearson correlations were run between the two probe type at each time period to examine WRCPM in each probe type each week. Correlation coefficients ranged from .93 to .40 with three correlation coefficients greater than 0.9, seven coefficients between .80 and .89, five between .70 and .79, two between .50 and .59, and one between .30 and .39. All correlations between the probe types at each time period were significant at the .05 level except the final week of data collection which had a correlation coefficient of .398 at a $p = 0.091$ significance level. Overall, these results indicate that the two types of measurement materials are measuring the same thing.

Slope Calculation

Ordinary Least Squares (OLS) regression analysis was conducted for each participant in each type of reading material, resulting in two slope values per participant. Each slope value represented the increase in the number of WRCPM per week using each type of probe for measurement. Slope values were also calculated for each grade level and the means and standard deviations for each probe type are reported in Table 7. Time (with week starting at 0) was used as the independent variable and WRCPM was the

dependent variable in each analysis. Data were collected two times per week and the mean WRCPM in each probe type was used in the analysis.

Table 7

Slope Descriptive Statistics by Type of Probe and Grade

		Open Court probes		AimsWeb probes	
Grade Level	N	Mean	SD	Mean	SD
1	26	1.48	0.81	1.49	0.67
2	21	1.26	0.69	0.92	0.98
3	17	0.14	0.6	0.41	0.5

To determine if differences in slope values were related to students' initial performance, correlations were computed between the slope and the intercept (determined by Ordinary Least Squares Regression) at each grade level. These correlations are displayed in Table 8. The only significant correlations between slope and intercept were for second grade AimsWeb and third grade Open Court. These results indicate that when measured with AimsWeb materials, the lower the second grade students' initial performance was, the faster their rate of growth was over the 19 weeks. Also, when measured with Open Court materials, the lower the third grade students' initial performance was, the faster their rate of growth was over the 19 weeks.

Table 8

Correlation Coefficients for Slope and Intercept by Grade

	AimsWeb		Open Court	
Grade	Correlation	Significance	Correlation	Significance
1	-.025	.902	.235	.248
2	-.443*	.044	-.112	.630
3	-.365	.149	-.508*	0038

*$p < .05$.

Analysis of Variance for Level

The first question to be answered by this study was *1) For elementary students, do curriculum-based probes result in the same number of words read correctly per minute as standardized probes developed outside of a curriculum?* To answer this question a 3 (grade) x 2 (probe type) repeated measures ANOVA using the three grade levels (first, second, and third) as a between group variable and the two probe types (Open Court and AimsWeb) as the within group variable was conducted with level as the dependent measure. Table 9 contains the ANOVA outcomes for main effect and interaction effects. This analysis indicated no significant interaction between grade and probe type, F (2, 61) = 0.30, p = .745. The level analysis did reveal a significant main effect for both grade, F (2, 61) = 17.36, p = .000 and probe type, F (1, 61) = 72.41, p = .000. These results mean that when averaged across grade levels, the mean WRCPM was significantly higher using AimsWeb measurement materials than Open Court materials. Also, when averaged across the two measurement probe types, the mean WRCPM of the participants in first, second, and third grades differed significantly. The AimsWeb probes yielded a mean of 35.35 WRCPM in first grade, 54.49 in second grade, and 71.84 in third grade. The Open Court materials yielded a mean of 27.18 WRCPM in first grade, 41.07 in second grade, and 66.22 in third grade. Post hoc analyses were conducted to determine where the grade level differences were.

Table 9

F-values for ANOVA Results of CBM in Reading Across Grade and Probe Types

Factor	Level	Slope
Grade	9.05*	22.72*
Probe Type	72.41*	0.03
Grade x Probe Type	0.30	2.19

*p < .01.

Post Hoc Procedures for Level

Tukey's Honestly Significant Difference (HSD) test was used in post hoc analysis multiple comparison procedure to clarify ANOVA results for the main effect for grade using level as the dependent measure. Table 10 outlines the post hoc comparison results. These results indicate that third grade students had the highest WRCPM and first grade students had the lowest WRCPM. However, the differences between third and second grade and between second and first grade were not significant (p>.05). The only significant difference between grade levels for WRCPM was for third grade students who had significantly more WRCPM than first grade students in the sample.

Table 10

Tukey's HSD Test Across Grade Levels for WRCPM Level

Grade Comparison	Mean Difference	Significance
Grade 3 - Grade 2	13.38	0.126
Grade 3 - Grade 1	27.27*	0.000
Grade 2 - Grade 1	13.88	0.066

*p < .01.

Because of the small sample size in this study and the resulting low power, effect sizes were calculated to further compare probe type materials at each grade level. The effect sizes were small at 0.30 for third grade and moderate for grades one (-0.48) and two (0.50). AIMSweb probes resulted in a higher level of WRCPM than Open Court

probes did at each grade level. Therefore, according to these analyses, curriculum-based probes (Open Court) do not result in the same WRCPM as standardized probes developed outside the curriculum (AIMSweb).

Analysis of Variance for Slope

The second and third questions to be answered by this study were

2) Do curriculum-based probes result in the same rate of increase over time in words read correctly per minute as standardized probes developed outside of a curriculum?

3) *Is there an interaction effect of type of probe and grade level for the rate of increase in words read correctly per minute over time?*

To answer these questions a 3 (grade) x 2 (probe type) repeated measures ANOVA using the three grade levels (first, second, and third) as a between group variable and the two probe types (Open Court and AimsWeb) as the within group variable were conducted with slope as the dependent measure. Table 9 contains the ANOVA outcomes for main effect and interaction effects.

There was no a significant interaction effect for grade and probe type, F (2, 61) = 2.19, p = .120. No significant main effect for probe type, F (2, 61) = 0.03, p = .856 was found. The slopes measured by AimsWeb materials at each grade level were not significantly different from those measured by Open Court materials. The mean slope for grade one measured with AimsWeb probes was 1.49, 0.92 for second, and 0.41 for third. The mean slopes measured with Open Court materials were 1.48 for first grade, 1.26 for second, and 0.14 for third. All slopes were smallest for grade three and largest for grade one. The slope analysis did reveal a significant main effect for grade, F (2, 61) = 22.72, p

= .000, meaning that when averaging across the two t probe types, the rate of reading growth is significantly different across grade levels. Again, post hoc analyses were conducted to determine where these differences lie.

Post Hoc Analysis for Slope

Tukey's Honestly Significant Difference (HSD) test was used in post hoc analysis multiple comparison procedure to clarify ANOVA results for grade level differences. Table 11 outlines the post hoc comparison results. These results indicate that students in first grade had significantly steeper slopes for WRCPM gained per week than third grade students and second grade students had significantly steeper slopes for WRCPM gained per week than third grade students in the sample. However, there was no significant difference between the growth in first grade and second grade.

Table 11

Tukey's HSD Test Across Grade Levels for WRCPM Slopes

Grade Comparison	Mean Difference	Significance
Grade 1 - Grade 2	0.39	0.059
Grade 1 - Grade 3	1.21*	0.000
Grade 2 - Grade 3	0.81*	0.000

*$p < .01$.

Effect sizes were calculated to further investigate differences in slope for the two probe types. The effect sizes were small at 0.01 for first grade and moderate for grade two (-0.40), and for grade three (0.49). The slope values were higher for AimsWeb materials in first and third grades. However, the negative effect size for second grade indicates that Open Court materials produced higher slope values in second grade. Therefore, although the slope values were not significantly different according to the ANOVA analysis, the rate of growth in WRCPM is higher measured by AimsWeb

materials than by Open Court materials for first and third graders and higher for Open Court in second grade.

The answer to question two based on these analyses is that the two types of probes result in similar growth rates at each grade level. No significant difference between was found between the growth rates in WRCPM for the two types of probes. For question three, these analyses showed that there was no significant interaction effect for grade and probe type despite the finding that one probe type did not yield faster growth in WRCPM at each grade level.

Growth Rate Analysis

Slope data were re-examined to answer question four, *4) Does the rate of increase over time in words read correctly per minute differ at different grade levels?* The mean slope at each grade level can be conceptualized as the mean weekly growth rate. These data for each grade level are displayed in Table 12. Results of the ANOVA using slope as the dependent measure indicated a significant main effect for grade level and post hoc analysis revealed significant differences in growth rates between first and third grade as well as second and third grade. Therefore, the growth rates of WRCPM do differ at different grade levels.

Table 12

Weekly Mean Growth in WRCPM by Grade

Grade	AimsWeb	Open Court
1	1.49	1.48
2	0.92	1.26
3	0.41	0.14

Chapter Five

Discussion

The current study was conducted to clarify and extend previous research on the comparibility of CBM ORF results using reading materials from outside the students' curriculum for repeated measurement over time. Specifically, this study evaluated the use of generic measurement materials for progress monitoring student reading growth over time. Additionally, this study investigated expected gains in words read correctly per minute over time at different grade levels.

The research questions addressed in this study were primarily focused on whether there were differences between curriculum-based reading probes and generic, standardized probes developed outside of a curriculum for measuring student progress. This chapter explains the results of this study in relation to previous research findings and discusses implications for future research and practice. Limitations of the study are discussed also.

Analysis of Level Data

The level analysis revealed a significant main effect for both grade and probe type. Significantly higher levels of WRCPM were found in AimsWeb probes when compared to Open Court probes. When compared by grade level, significant differences were found between first and third grade only. Research has consistently shown differences in the level of reading performance of students measured with probes

developed from different curricula (e.g., Hintze, Shapiro, & Lutz, 1994; Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; and Powell-Smith & Bradley-Klug, 2001).

Several studies compared literature-based to basal-based probes (i.e., Hintze, Shapiro, & Lutz, 1994 and Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998). Bradley-Klug, Shapiro, Lutz, and DuPaul (1998) found that although both probe types displayed sensitivity, mean performance using basal probes was significantly higher than mean performance using literature-based probes. The literature-based probes may have underestimated the level of student performance. In addition, the literature-based probes may have contained a greater percentage of unfamiliar vocabulary than the basal probes.

The current study results most resembled the research from Powell-Smith and Bradley-Klug (2001) which compared generic (no specific curriculum) and basal probes. As in the current study, the participants read significantly more words correctly using generic measurement probes than using basal probes.

This difference between the Powell-Smith and Bradley-Klug (2001) and the current study could be due to differences in readability between the two types of probes. Generic probes such as AimsWeb are controlled for strict readability for each passage (Howe & Shinn, 2001), but basal probes such as Open Court (1997) are controlled for phonetical concepts also with reading passages being focused on the concept being covered for that section of the curriculum.

Overall, the mean words read correct per minute (WRCPM) were comparable to results of several other studies although a bit lower (e.g., Hintze, Shapiro, & Lutz, 1994; Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; Powell-Smith & Bradley-Klug, 2001). The students in this study were reading a mean of 27.18, 41.07, and 66.22 WRCPM in

first, second, and third grade, respectively in Open Court probes and 35.35, 54.49, and 71.84 in AimsWeb probes. Bradley-Klug et al. found that second grade students read a mean of 58.64 and 54.50 words correct per minute in basal and literature based material, respectively. Powell-Smith and Bradley-Klug found that second grade students read a mean of 38.3 and 50.0 words correct per minute in basal and literature-based material respectively.

Despite some similarity in mean WRCPM among studies, benchmarks for successful reading have increased in recent years. The mean WRCPM were low compared to Florida state benchmarks as well as reading benchmarks established by the authors of DIBELS (Dynamic Indicators of Basic Early Literacy Skills). This sample, as a group, were performing substantially below Florida established benchmarks of 40, 90, and 110 at the end of first, second, and third grade respectively (Florida Center for Reading Research, 2004). In addition, slope data were low particularly for grades two and three leading to the prediction that these students will fall further and further behind what is expected of them in school curriculums and on high stakes tests. Research has established measures of oral reading fluency as valid and reliable predictors of both overall reading outcomes and performance on high stakes tests (Good, Simmons, & Kame'enui, 2001). The mean WRCPM for each grade level of the current study fall into the "moderately below grade level and in need of additional intervention" category of the Florida benchmarks for oral reading fluency (Florida Center for Reading Research, 2004). In addition, these data are means and many of the students in the sample had significantly lower levels of performance. These data suggest that many of the participating students were in need of intensive remediation to increase growth rates and

bridge the gap between their performance and the established indicators for success in school.

Analysis of Slope Data

The slope analysis did not reveal a significant main effect for probe type. Slopes were not significantly different based on the type of probe that was used to monitor progress. Both AimsWeb and Open Court probes were similarly sensitive to reading growth over time. Research on varying measurement material for CBM ORF has shown similar results with comparably linear and stable growth over time (Hintze & Shapiro, 1997; Bradley-Klug et al., 1998; Powell-Smith & Bradley-Klug, 2001).

However, not all studies found equal sensitivity to growth over time for CBM ORF measurement material from different sources. In Hintze, Shapiro, and Lutz (1994), regardless of the reading series in which the students were taught, performance declined when students were monitored using the literature-based series and performance increased when monitored with the traditional basal series. Therefore, the use of probes developed from the traditional basal series for progress monitoring showed greater sensitivity to growth over time than the use of literature-based probes.

Also King, Bradley-Klug, Knoff, and Powell-Smith (2002) found varying sensitivity to reading growth when different measurement materials were used for progress monitoring. This research showed a significant main effect of curriculum on mean regression slopes. The slope of improvement averaged across reading levels was highest for the Accelerated Reader Program authentic literature probes and lowest for the Scott Foresman literature-based probes. There was also a main effect of instructional reading level on the oral reading rate mean regression slopes. Finally, King et al. found

that the mean regression slopes differed significantly as a function of instructional level when probes from either the direct instruction or the literature-based curricula were used. However, probes developed from the authentic literature curriculum worked equally as well for students at all three instructional levels.

Methodological differences may contribute to the disparity of findings across studies. The Hintze et al. (1994) study used reading probes that were developed only from third grade material in each of the reading series. The sample in the King et al. study consisted solely of students receiving special education services. In addition measurement material and group placement were determined by instructional level rather than grade level.

Slope Differences by Grade

The slope analysis revealed a significant main effect for grade. Further examination of slope by grade level revealed higher slopes for first and third graders in AimsWeb measurement material, but higher slopes for second graders in Open Court material. Although the slopes were not significantly different at any grade level, one type of measurement material was not more sensitive to growth over time for each grade level.

Hintze and Shapiro (1997) found similar results with larger growth rates for second, third, and fourth grade students when monitored in literature-based probes. However, fifth grade students showed higher growth rates when monitored with basal probes. Studies have not shown a consistent pattern for which probe types show more growth at which grade level.

In the current study, results showed smaller slopes as a function of increasing grade level. Slopes were significantly greater for grades one and two than for grade

three. Research tends to show that the average weekly gains in words read correctly per minute decrease with increasing grade level (Fuchs, Fuchs, Hamlet, Walz, & Germann, 1993; Marston & Magnusson, 1985; Marston and Tindal, 1995). However, unlike other research, King, Bradley-Klug, Knoff, and Powell-Smith (2002) did not find that students in lower grade instructional levels displayed higher oral reading growth rates with a sample of special education students. Instead, the students at the first grade instructional level exhibited the lowest growth rates when reading rates across curricula were averaged. Students at the second grade reading level had the highest rate of progress. In addition, Hintze and Shapiro (1997) monitored reading progress of general education students in basal-based and literature-based curricula. Results revealed increasing slopes as a function of increasing grade levels. The only exception to this finding was higher slopes for grade four compared to grade five when monitored with literature-based probes.

One explanation for this disparity in findings is that the students were grouped by instructional level in the King et al. study and by grade level in the Fuchs et al. (1993) and Marston and Tindal (1995) studies. This grouping information means that in the King et al. study a fourth grade student could be monitored with first grade reading material and be compared to fourth grade students monitored in fourth grade reading material. Other research has compared student reading progress only by grade level whether progress was monitored using grade level, instructional level, or above grade level materials.

Expected Reading Growth

Another purpose of this study was to examine expected reading growth rate at each grade level in each type of probe. AimsWeb and Open Court results for first grade were very similar with a growth of 1.49 and 1.48 words per minute respectively. In second grade weekly rates of change were 0.92 for AimsWeb and 1.26 for Open Court. The third grade AimsWeb slope in WRCPM was 0.41 and 0.14 for Open Court. These slopes are lower than those found in other studies of general education students' reading gains (Deno, Fuchs, Marston, & Shin, 2001; Fuchs et. al., 1993; Marston & Magnusson, 1985). Although the current findings for growth over time are lower than other results for general education students, they more closely resemble results for students receiving special education services. Marston and Magnusson found greater growth for first, second and third grade students than the current study with growths of 1.6, 1.4, and 1.9, respectively. However, current study results were greater than results of the Deno et al. study for students in special education in first grade (0.83) and second grade (0.57), but lower than the third grade results (0.58). Similarly, current results for weekly gain were higher for first (0.0007) and third grades (0.003), but comparable for second grade (0.931).

Some of these differences across these studies may be due to the number of data collection points in each study. In the current study, students were monitored twice a week for 17 weeks across a 19 week period. However data collection periods varied widely across previous studies. For example, Deno et al. (2001) collected data only three times in a year, Marston and Magnusson (1985) collected data once a week for 16 weeks, and King et al. (2002) collected data twice a week for eight weeks.

Differences also may be due to variability in participant populations. For instance, data for both special education and general education students were combined in the Fuchs et al. (1993) study, whereas Deno et al. (2001) and Marston and Magnusson (1985) separated the two groups to report results. The current study monitored general education students who were primarily African-American (95%) as opposed to other studies with little minority representation. For example, Bradley-Klug et al. (1998) had a 5% minority with 1% African American sample. Powell-Smith and Bradley-Klug (2001) had 24% minority participants, but no African-Americans were in their sample. Hintze et al. (1994) used two schools in analysis with school A having 8% minority with no African American participants and school B having 25% minority with 12.5% African American participants.

Another population variable that may also be a factor is the behavior problems evident in the school from which the current sample was taken. Most students at the charter school had frequent discipline problems in other educational settings. Research has shown a correlation between behavior problems and difficulty reading. For example, Vaughn, Hogan, Lancelotta, and Shapiro (1992) compared six groups of kindergarten students with and without behavior problems, which included attention difficulties, conduct problems, and anxiety. The students without behavior problems demonstrated higher reading achievement scores than did the groups with reported behavior problems as indicated by the Revised Behavior Problem Checklist, peer ratings of social acceptance, teacher ratings of social skills, and self-ratings. Also, Kulekowskis (1996) examined reading achievement related to behavior problems for second grade students. Students identified by their teachers as having "troubling" behavior had significantly

lower reading achievement scores on the Metropolitan Achievement Test-Seven than students who were classified as "non-troubling."

Another possible reason for the differences in growth rates across studies may be the level of the material used for monitoring. In this study, grade level material was used for monitoring regardless of the students' reading levels. The only requirement was that the students read at least 10 words per minute in the grade level materials.

Results of a Marston and Magnusson (1985) study showed the greatest slope for material in which the students read 11 to 20 words correct per minute for first and second graders and 41 to 50 words read correctly per minute for third through sixth graders. In 1988, the same researchers found that between 10 and 30 words correct per minute for first and second grade students and between 30 and 60 words for third through sixth graders yielded the highest slopes.

Shinn, Gleason, and Tindal (1989) found no significant differences in slope between easier passages (one level below instructional level) and more difficult passages (one level above instructional level). Although the results indicated that the difficulty of the reading passage did not affect student progress, the researchers suggested that the failure to find differences in the slopes may have been due to the small sample size (30 students) and low power.

Finally, Hintze, Daly, and Shapiro (1998) found higher slopes in grade level compared to goal level material for first and second grade students. There were no significant differences in slopes for third and fourth grade students.

Correlation Between Initial Level of Performance and Rate of Growth

Educators often assume that students who perform at low levels also display

slower growth, however the results of this study suggest that even when a student has low initial skills it does not mean that they will have a slower rate of progress. This finding is consistent with findings of Powell-Smith, Shinn, Stoner, and Good (2000). King et al. (2002) also found that when measured with Accelerated Reader Program materials, students who displayed the lowest level of performance had the highest rate of growth over time.

Limitations

This study utilized a pre-existing data set that prohibited alterations to much of the methodology (e.g., participant selection). Given these circumstances, the results of this study have to be interpreted in light of several limitations. First, this study utilized only 64 participants, which indicates group sizes of 26, 21, and 17 students. A larger data set might yield different results. Second, the participants were drawn from only one school. This school may or may not be similar to other schools in educational standards, teaching methodologies, and student population. These unknown variables will require caution in generalizing the results of this study to other schools. Third, utilizing participants from only three grade levels (i.e., first, second, and third) may limit the external validity of the results, such that findings may only be generalized to students in first through third grades. Fourth, the participants were 64% male, 95% African American, and many had behavioral problems, which does not accurately reflect the ethnic composition of the county, state, or country. The generalizability of results may be limited by the overrepresentation of African-American and male students. However, the representation of African-Americans in this study is in sharp contrast to other studies in the area of reading assessment with very few African-American participants.

Therefore, generalization to African-American populations may be improved over compared to other similar studies. There are also limited data on interscorer accuracy and teachers collected only data for their own classes. Differences in scoring among teachers and situational factors in administration (different teachers administering passages, different locations within classrooms, different times of day for testing, knowledge of timing) may have affected results of the analyses. Previous research has found that student reading rate may be affected by who administers the probes, where the probes are administered, and if the student knows he or she is being timed (Derr-Minneci & Shapiro, 1992). Additionally, probes were not counterbalanced for data collection. Data were usually collected for Open Court probes followed by AimsWeb probes with no random order. Also, AimsWeb and Open Court probes were each done in the same order for each student, which may constitute a threat to internal validity. Lastly, this study measured using grade level material regardless of reading level. This choice may have resulted in ceiling effects for high level readers. The greatest growth in reading has been shown when measured just at or just beyond instructional level. Given these limitations, generalizability may be limited beyond the immediate participant sample and further research may be needed to confirm the results of this study.

Implications for Future Research

The results of the current study could be further validated through replication. In addition, similar research with various minority groups and students with various behavior issues and disabilities could shed light on any differences between groups related to measurement using various materials. Future research could also serve to distinguish if differences in the growth rates in the current study and previous research

with predominately Caucasian populations are due to the measurement itself (e.g., time factor), types of measurement material, and/or low levels of reading growth in the curriculum in which the students are taught. A similar study could also highlight differences in measurement materials and reading growth for students with behavioral concerns to further investigate what impact the behavioral concerns have on reading level and rate of reading growth.

Future research may also attend to the differences in level among differing measurement materials to determine if level benchmarks for successful reading outcomes and high stakes testing are different with differing probe types. It may be beneficial to know the benchmarks that are specific to the measurement material educators use. Looking at the differences in benchmarks may be more pragmatic for literature and basal based materials than the generic probes that are more strictly controlled for readability and seemingly similar. Similarly, the large-scale movement toward using generic probes may explain higher benchmark expectations than previously used instructional placement standards. This is because this research showed that levels of WRCPM were higher when generic probes were used for measurement rather than curricular probes. Future research could examine whether expectations for students have increased in this area or if the change in widely used sources of measurement material has driven the increase in benchmark standards.

Implications for Educational Practice

The use of generic, not curricular probes was shown to be equally as sensitive to growth over time as basal probes. The generic probes may be more efficient for use in an educational setting because they are pre-made and often are accompanied by technology

for graphing and organizing progress monitoring results. As technology becomes increasingly vital in education, generic probes may become more attractive to educators for abundant comparison with other educators and ease of use.

The disparity in findings between this study and other studies with primarily Caucasian participants sheds light on the need to monitor the progress of various minority groups within a school to detect problems with the academic culture of a school or the inequity of the education therein. This need becomes apparent with Adequate Yearly Progress requirements under No Child Left Behind (2001) in which all groups are expected to show the required progress, not just the overall school. CBM ORF presents as one way to keep track of progress within different groups and indicate where changes are needed.

Conclusion

This study contributed to the present CBM ORF literature base in several ways. First, this study supports the comparability of two sources of measurement material (i.e., AimsWeb, Open Court) for CBM ORF progress monitoring that had not been previously examined and found that they are approximately equally sensitive to growth over time. Second, by collecting progress monitoring data for a longer period of time (i.e., 19 weeks) than other studies examining the relationship between reading growth over time and the material used to measure it (i.e., five to ten weeks), this study provided a more complete picture of reading growth over time. The results of this research helped to clear up the disparity in findings. Third, by examining differences in expected growth rates due to differences in measurement materials, this study added to the research on academic growth standards for students especially as it relates to African-American populations. In addition, this research has practical impact in informing decisions about the choice of progress monitoring materials and assisting in the development of appropriate goals for improvement in ORF with the establishment of academic growth standards in varying curricula. Finally, the sample consisted primarily of students with behavioral difficulties who were primarily African-American. The use of this sample provided insight and preliminary data regarding differences among groups of students who differ by ethnicity and behavior utilizing measurement from different sources for CBM ORF progress monitoring.

References

Bradley-Klug, K. L., Shapiro, E. S., Lutz, J. G., DuPaul, G. J. (1998). Evaluation of oral reading rate as a curriculum-based measure within literature-based curriculum. *Journal of School Psychology, 36*, 183-197.

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

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 1-17). New York: Guilford Press.

Deno, S. L., Deno, D., Marston, D., & Marston, D. (1987). *Test of reading fluency: Measures for screening and progress monitoring.* Minneapolis, MN: Children's Educational Services.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-523.

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

Dunn, L. M., & Markwardt, F. C. (1970). *Peabody Individual Achievement Te*st. Circle Pines, MN: American Guidance Services.

Engelmann, S., & Bruner, E. C. (1995). *Reading Mastery* (Rainbow ed.). Worthington, OH: SRA Macmillan/McGraw-Hill.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*, 232-243.

Fuchs, L., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15-24.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1982). Reliability and validity of curriculum-based Informal reading inventories. *Reading Research Quarterly, 18(1)*, 6-26

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Waltz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27-48.

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

Fuchs, L.S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum-based reading assessment. *Diagnostique, 9,* 191-207.

Good, R.H., Simmons, D.C., & Kame'enui, E. (2001). The importance and decision making comparibility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. Scientific Studies of Reading, 5, 257-288.

Hartman, J. M., & Fuller, M. L. (1997). The development of curriculum-based measurement norms in literature-based classrooms. *Journal of School Psychology, 35(4),* p 377-389.

Hintze, J. M., Daly III, E. J., & Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27,* 433-445.

Hintze, J. M, Shapiro, E. S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*, 351-375.

Hintze, J. M, Shapiro, E. S., Conte, K. L., & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26*, 535-553.

Hintze, J. M., Shapiro, E. M., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education, 28*, 188-202.

Howe, K. B., & Shinn, M. M. (2001). *Standard reading assessment passages for use in general outcome measurement: A manual describing development and technical features.* Eden Prairie, MN: Edformation.

Karlsen, B. L., Madden, R., & Garner, E. (1975). *Stanford Diagnostic Reading Test* (2nd ed.). San Antonio, TX: Psychological Corporation.

King, M. L., Bradley-Klug, K. L., Knoff, H. M., & Powell-Smith, K. A. (2002). *Tracking the reading progress of students with specific learning disabilities: A comparison of direct instruction, literature-based, and authentic reading curriculum-based measures.* Unpublished manuscript, University of South Florida.

Koslin, B. L., Koslin, S., Zeno, S. M., & Ivens, S. H. (1989). *The degrees of reading power test: Primary and standard forms.* Brewster, NY: Touchstone Applied Science Associates.

Kulekowskis, J. (1996). *The effect of student behavior on the reading achievement of second grade students.* Eric document ED397398

Madden, R., Gardner, E. R., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1973). *Stanford Achievement Test*. New York: Harcourt, Brace, Jovanovich.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*, 266-276.

Marston, D., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137-172). Washington, DC: National Association of School Psychologists.

Marston, D., & Tindal, G. (1995). Performance monitoring. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology III* (pp. 597-607). Washington, DC: National Association of School Psychologists.

Naslund, R. A., Thorpe, L. P., & Lefever, D. W. (1978). *SRA Achievement Series.* Chicago: Science Research Associates.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development.

Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the "C" in CBM: Does it really matter if curriculum-based measurement reading probes are curriculum-based? *Psychology in the Schools, 38*, 299-312.

Powell-Smith, K. A., & Stewart, L. H. (1998). The use of curriculum-based measurement in the reintegration of students with mild disabilities. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 254-297). New York: Guilford Press.

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

Shinn, M., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*, 31-38.

Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 23,* 223-233.

Shinn, M. R., Good, R. H., III, Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459-479.

Sum, A. (1999) Literacy in the Labor Force: Results from the National Adult Literacy Survey Publication number NCES1999470.

Stenner, A. J., Smith, M., & Burdick, D. S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement, 20,* 305-315.

Tindal, G. (1988). Curriculum-based measurement. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative Educational Delivery Systems: Enhancing Instructional Options for all Students* (pp. 111-135). Washington, DC: National Association of School Psychologists.

Tindal, G., Germann, G., & Deno, S. L. (1983). *Descriptive research on the Pine County norms: A compilation of findings* (Res. Rep. No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Res. Rep. No. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tucker, J. (1987). Curriculum-based assessment is no fad. *The Collaborative Educator, 1,* 4-10.

Vaughn, S., Hogan, A., Lancelotta, G., & Shapiro, S. (1992). Subgroups of children with severe and mild behavior problems: Social competence and reading achievement. *Journal of Clinical Child Psychology. 21*.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Services.

Yell, M., Deno, S. L., & Marston, D. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique, 18(1)*, 99-112.