Curriculum Based Measurements as Predictors of the Missouri Assessment Program Assessments in Grades 3-5

Kristen Annette Childers
B.S., William Jewell College, 2000
M.A., University of Missouri-Kansas City, 2004

Submitted to the Graduate Department and Faculty
of the School of Education of Baker University
in partial fulfillment of the requirements for the degree

Doctor of Education
in
Educational Leadership

November 18, 2013

Dissertation Committee

_____

Major Advisor

_____

_____

_____

Abstract

This quantitative study included third, fourth, and fifth grade students from the North Kansas City Schools, a school district in suburban Kansas City, Missouri. The purpose of this study was to determine if Curriculum Based Measurement (CBM) scores in reading and math and demographic variables of gender, race, and socioeconomic status (SES), were predictors of student scores on the Missouri Assessment Program (MAP) assessments in communication arts and math. The stepwise multiple regression models revealed that CBM scores were strong predictors of MAP scores. However, the demographic variables gender and race were only included in models as predictors of communication arts MAP scores for third grade. SES was a predictor across all grade levels and in both subject areas, except for MAP math scores for fifth grade. It is recommended to utilize CBM as a formative assessment to provide teachers with data to guide their instruction.

Dedication

This dissertation is dedicated to my family. First, to my mother, my cheerleader, my confidant, and my best friend: Marilyn Kuhlman. She is the woman who taught me to reach for the stars and never give up on my goals. I was lucky to have her as a strong, independent female role model. Secondly, to my father: Dan Kuhlman. Always ready with a solution and a story to help me get though the rough times along this journey. Finally, to my loving husband: Scott Childers. I was blessed when I met you. Thank you for your encouragement, love, and help through the time spent on this project. Thank you for understanding that quality time sometimes meant just being in the same room with me as I worked on the laptop. I appreciate you and your continued support.

Acknowledgements

The importance of an education was instilled in me from an early age. My parents, both educators, motivated me as a learner. Through the years, under their watchful eyes, I encountered many wonderful teachers. I would like to thank educators, the world over, for helping students like me in their pursuit of an education. Specifically, I would like to acknowledge my professors and advisors from Baker University. Dr. Susan Rogers supported and encouraged me from the start of my degree program. I am ever grateful for her knowledge, input, and feedback. Additionally, I would like to acknowledge the work of Peg Waterman and Katie Hole. Both ladies graciously shared their statistical expertise to assist me with this work. Finally, I would like to recognize Dr. Lisa Friesen and Dr. Brett Knappe for their work on my dissertation committee. Thank you for your time and support.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

School districts are accountable for the learning achieved within their schools at the national level by the U.S. Department of Education. With the passage of the No Child Left Behind Act of 2001 (NCLB), school districts are required to report assessment data for students in third through eighth grades, and one additional grade at the high school level, in the areas of language arts and mathematics (Altshuler & Schmautz, 2006). Each state has become responsible for creating both curriculum standards and the assessments utilized to determine student proficiency, as mandated by NCLB (Myers, 2008).

There are two widely accepted categories of assessment: summative and formative. Summative assessments show gains that students have made over time. Summative assessments are "those assessments that happen after learning is supposed to have occurred to determine if it did" (Stiggins, Arter, Chappuis, & Chappuis, 2006, p. 31). Therefore, districts report summative assessment data to the state to show the progress that students have made towards yearly goals. However, in the case of standardized summative state assessments, teachers do not have data available for analysis until students have moved to the next grade.

Formative assessments, in contrast, provide immediate data detailing students' progress. Teachers utilize the formative assessment data in order "to diagnose student needs, plan our next steps of instruction, provide students with feedback they can use to improve the quality of their work and help students see and feel in control of their journey to success" (Stiggins et al., 2006, p. 31). Curriculum Based Measurements

(CBMs) are formative assessments administered within school settings to monitor student progress. Teachers can utilize CBM data throughout the year to "monitor student progress toward long-term goals" (Steckler, 2006, pp. 92-93).

Background of the Study

The state of Missouri has 522 school districts and 33 charter schools that educate over 917,000 students who are enrolled in pre-kindergarten through grade 12 (Department of Elementary and Secondary Education [DESE], 2011c). Schools in Missouri are responsible for complying with Missouri Senate Bill 380 (the Outstanding Schools Act) and NCLB mandates. The Missouri School Improvement Program (MSIP) is responsible for reviewing and accrediting the school districts of Missouri to determine if they are following the protocols set forth by state and national boards (DESE, 2010b). In order to show active participation at each level, all students in designated grades and curriculum areas are mandated to take an assessment to demonstrate their overall growth (DESE, 2006; State of Missouri, 2009). The Missouri Assessment Program (MAP), a series of standardized tests developed to fulfill the requirements of the Outstanding Schools Act of 1993, assesses the knowledge that students were to acquire throughout their schooling (University Academy, 2009). The MAP is a summative assessment that is a measure of knowledge gained, which is administered to meet the mandates of NCLB (M. Muenks, personal communication, June 28, 2010).

The North Kansas City School District 74 (NKCS) is located in the southern area of suburban Clay County, Missouri, and includes an early childhood program, 21 elementary schools, five middle schools, four high schools, and several alternative education programs. During the 2010-2011 school year, 18,523 students were enrolled in

the district schools (DESE, 2011a). This study's focus was on elementary grade levels in which the MAP assessments are administered. This included 1,415 students enrolled in the third grade, 1,461 students enrolled in the fourth grade, and 1,432 students enrolled in the fifth grade (NKCS, 2010a). Table 1 summarizes the disaggregated demographic data for grades 3-5 in the NKCS during the 2010-2011 school year.

Table 1

2010-2011 North Kansas City Student Demographic Data

Demographic	Third Grade	Fourth Grade	Fifth Grade
Gender			
Male	732	711	774
Female	683	750	658
Race			
White	858	969	928
Non-White	557	492	504
Socioeconomic Status			
Free or Reduced	847	844	820
Full Pay	740	771	782
Total Students	1,415	1,461	1,432

Note. Adapted from "Core Data Report," by NKCS, 2009.

The NKCS has developed an extensive Assessment Plan that includes participating in the annual state mandated MAP (NKCS, 2009). Communication arts and math are assessed in grades 3 through 8 (DESE, 2010a). The MAP assesses students through constructed response, multiple choice, and performance events (DESE, 2009). A

performance event assessment requires students to perform a real-world task to demonstrate mastery of the curriculum (Zane, 2009).

Unlike the MAP, CBMs are utilized at the district level. The NKCS formally started to utilize the CBM as a source of data during the 2005-2006 school year with its adoption into the District's Assessment Plan (Smith, 2007). The original reasoning for the use of CBM assessments was for benchmarking student achievement levels and monitoring for special education qualification (Keller-Margulis, Shapiro, & Hintze, 2008). The NKCS changed from using a formula for qualifying students into the special education system, the IQ/Achievement Discrepancy Model, to the Response to Intervention (RtI) Model during the 2005-2006 school year (Smith, 2007). Therefore, teachers and administrators abandoned using the CBMs as originally purposed and have since used them as a guiding assessment for instruction.

One component within the RtI framework allows teachers to monitor students as they participate in research-based interventions. If students are able to improve because of the interventions, then they are responding positively to the intervention and do not need special education services. However, if a student does not respond to interventions, then recommendations for possible special education assistance is proposed. Prior to the use of the RtI framework, students were given a battery of standardized assessments in order to determine eligibility for special education services. However, within this process, students who had to be referred by teachers and at-risk students continued to fall behind academically. The CBM benchmarking allows for an analysis of achievement three times a year: fall, winter, and spring. Teachers utilize the data to determine at-risk students and to determine growth between the benchmarking periods for all students

(Deno, 1986). Thus, the CBM assessments help teachers to monitor students and change instruction to meet the needs of their students within the general education classroom.

Research and development of CBMs began in the mid-1970s at the University of Minnesota with the work of Deno and a group of graduate students in the educational psychology field. Deno and his team were working to discover a grouping of tasks that would assist educators in measuring student improvement. Deno (1985) stated that CBM data must meet the following standards before being formally presented to teachers:

1. Reliable and valid if the results of their use were to be accepted as evidence regarding student achievement and the basis for making instructional decisions.
2. Simple and efficient if teachers were going to use them, or teach others to use them, to frequently monitor student achievement.
3. Easily understood so that the results could be clearly and correctly communicated to parents, teachers, and students.
4. Inexpensive since multiple forms were to be required for repeated measurement. (p. 221)

For items 2-4, Deno (1985) shared that CBMs must have the listed characteristics due to the frequency of these assessments. If the guidelines are not followed, teachers may not understand the administration, scores may be harder to understand in relation to grade level curriculum, and cost could inhibit widespread use. With these criteria met, the CBM data allows teachers to make decisions regarding their students' current levels of achievement (Deno, 1986). In the NKCS, teachers follow these guidelines for CBM usage.

Statement of the Problem

Formative and summative assessments are vital for data analysis relating to student achievement. The NKCS continually works to prepare students for the summative MAP assessment. The MAP provides teachers with an overview of the learning that took place within the classroom during that calendar year. However, teachers were not able to use the results to guide their curriculum decisions and monitor student growth through the use of annual summative assessments (International Reading Association, 1999). Thus, there needs to be a mix of summative assessments and formative assessments for optimal progress monitoring of students (Gunzenhauser, 2003; Semas, 2001). This can be achieved through a comprehensive assessment plan, which includes both formative and summative assessments (Burke, 2010).

The district periodically assesses students utilizing the formative CBMs in order to determine if students are making progress towards mastery of ongoing achievement goals. As of the 2010-2011 school year, the NKCS and other districts utilizing the CBMs do not know whether there is a relationship between students' ongoing CBM assessment scores and the students' scores on the spring MAP assessments in communication arts and math. It is important to know if there is a relationship between the two assessments in order to justify the usage of the CBMs as a source of predictive assessment data.

Purpose Statement

The purpose of this study was to determine the extent of the relationship between scores in reading and math on the CBM and scores in communication arts and math on the MAP assessments for third, fourth, and fifth grade students enrolled in the NKCS. An additional purpose was to determine if demographic data (gender,

socioeconomic status [SES], and race) and the CBM assessment scores are predictors of MAP scores.

Significance of the Study

The results of this study could provide valuable information to school districts concerning the importance of CBM assessments relative to student success on the MAP. Districts would then be able to determine whether the classroom time taken to administer the CBM was worth the investment. The administration of a CBM is a district requirement for the grade levels included in this study; however, unless a relationship between achievement on the CBM and the MAP is determined, the need to continue administration of the CBM versus other formative assessments is questionable.

Secondly, the analysis of demographic group data could lead to the discovery of new information concerning student achievement within NKCS. Subgroups may perform differently on assessments. Therefore, it is important to determine how subgroups perform on assessments in order to find one that is the best choice for assessing all subgroups.

Delimitations

Delimitations are "boundaries set by the researcher on the purpose and scope of the study" (Lunenburg & Irby, 2008, p. 134). This study involved delimitations that centered on the one school district chosen for the study.

1. The sample was delimited to 3rd, 4th, and 5th grade students in the selected district.
2. The demographic variables were limited to include gender, SES, and race.

3. Outcomes were limited to including student achievement only in mathematics and reading on the CBM and communication arts and mathematics on the MAP.
4. The sample only included students who participated in each of the fall, winter, and spring CBM assessments and the spring MAP assessment.

Assumptions

According to Lunenburg and Irby (2008), "assumptions are postulates, premises, and propositions that are accepted as operational for purposes of the research" (p. 135). This study included the following assumptions:

1. Teachers administered the CBM assessments in a standardized manner.
2. Teachers scored the CBM assessments for their students in a standardized manner.
3. Individual school personnel entered their student data into the Pearson PsychCorp AIMSweb program accurately.
4. Teachers administered the MAP assessments in a standardized manner.
5. State personnel scored the MAP assessments in a standardized manner.
6. Handling of the materials for the MAP assessments was performed in an ethical and legal manner, following state guidelines.
7. Students put forth their best effort on all given assessments.

Research Questions

This study focused on finding the best combination of variables for predicting student achievement on the MAP assessment. The following research questions provide the basis for this study.

RQ1: What combination of variables (fall Oral Reading Fluency CBM score, winter Oral Reading Fluency CBM score, spring Oral Reading Fluency CBM score, gender, SES, and race) best predicts student scores on the third grade MAP assessment in communication arts?

RQ2: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in communication arts?

RQ3: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in communication arts?

RQ4: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the third grade MAP assessment in math?

RQ5: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in math?

RQ6: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in math?

Definition of Terms

Assessments terminology and legislation can include similar concepts with different titles and acronyms. Therefore, the following section includes terminology

definitions centered on the theory and application components of assessment measures included in this study.

Constructed response. A constructed response test item requires a student to supply information to solve a problem (DESE, 2009).

Curriculum Based Measurement (CBM). CBM is a standardized tool used to measure student success within the core education skills that are taught within the educational system (Deno, Lembke, & Anderson, 2005). Long-term objective goals utilize the CBM assessments (for reading, mathematics, spelling, and written expression curriculum areas) in order to get quick assessment data of student achievement (Hintze & Christ, 2004; Hintze & Silberglitt, 2005). Standardized scoring techniques provide for the possibility of student score comparisons (Hintze & Silberglitt, 2005).

Depth of Knowledge (DOK). Depth of Knowledge contains four levels of questioning. The levels include recall, skill/concept, systematic thinking, and extending thinking (Webb, 2006). When the DOK level of questions align with DOK curriculum levels, academic rigor and expectations are maintained (Venet, 2009).

Fluency. Meyer and Felton (1999) defined fluency as "the ability to read connected text rapidly, smoothly, effortlessly, and automatically with little conscious attention to the mechanics of reading, such as decoding" (p. 284).

Formative assessment. Formative assessment is an assessment strategy that teachers can utilize, during the school year, to influence instructional choices by teachers. In addition, teachers utilize the data to provide students with feedback that will help them to increase achievement (Marzano, 2009).

Grade Level Expectation (GLE). GLEs refer to the documents created by the Department of Elementary and Secondary Education of Missouri that show the curricular expectations for each grade level. These were created in order to align with the Missouri State "Show-Me" Standards upon which the MAP assessment is based (DESE, 2005).

Math CBM (M-CBM). The M-CBM is a CBM probe to measure a student's ability in the computational areas of addition, subtraction, multiplication, and division. Students have a time limit of two minutes to work and they are assessed on the total correct digits written with corresponding problems (Shinn, 2004).

Maze CBM. The Maze CBM is a multiple-choice assessment that students complete while silently reading a passage. According to Pearson Education, Inc. (2008), students are given a passage to read in which the "first sentence of a 150-400 word passage is in its entirety. Thereafter, every seventh word is replaced with three words inside parentheses. One of the words is the exact one from the original passage" (para. 2). The Maze CBM is also known as maze task or multiple-choice cloze assessment (Wiley & Deno, 2005).

Missouri Assessment Program (MAP). The MAP is a yearly assessment that measures the students of Missouri and their mastery of the Missouri State "Show-Me" Standards for specific grades and content areas. The assessment contains sections of the *TerraNova* survey, a national norm-referenced test, and student comparisons to peer groups across the country take place (DESE, 2011b).

Multiple-choice items. Multiple-choice items are a response format in which respondents are to select the best possible answer (or answers) out of the choices from a

list. The utilization of this type of standardized testing item is common due to quick grading techniques (Stiggins et al., 2006).

Oral Reading Fluency Curriculum Based Measurement (R-CBM). The R-CBM is a teacher-administered assessment to determine a student's reading ability. In order to administer this assessment, students are required to read aloud for 1 minute. The number of words read correctly and errors are counted and scored. Standardization occurs when teachers utilize the reading passages within the Pearson PsychCorp's AIMSweb system. The strength of the R-CBM assessment, according to Hintze and Silberglitt (2005), is "its ability to serve as a broad signal of the multifaceted construct of reading and its ability to index student performance across a variety of contexts" (p. 374).

Oral Reading Fluency CBM errors. An error for the R-CBM assessment is any of the following: "mispronunciation of the word or substitution, omission, [and/or] 3 second pauses or struggles" (Shinn & Shinn, 2002b, p. 13).

Pearson PsychCorp AIMSweb. AIMSweb is a benchmark and progress monitoring system, which provides teachers with standardized assessments to utilize within their classroom (Pearson PsychCorp, 2010a). Additionally, the system allows the input of CBM scores into the system for retrieval at a later time.

Socioeconomic status (SES). SES determines the student's eligibility for a free and reduced lunch status, as set by the U.S. Department of Agriculture time-line of July 1, 2010 through June 30, 2011 (U.S. Department of Agriculture, 2011a). However, during the year 2010-2011, the requirements were the same as the year prior (U.S. Department of Agriculture, 2010). The eligibility requirements are determined by the number of people living in a household and the annual income of the household (U.S.

Department of Agriculture, 2011b). If students do not qualify for the free and reduced lunch status, their status is full pay (see Appendix A).

Summative assessment. The purpose of a summative assessment is to measure what a student has or has not learned within a unit or school year (Burke, 2010). It is a final judgment to show mastery level of taught objectives.

Words read correctly. In terms of the R-CBM, the words read correctly are words pronounced correctly in context and self-corrected errors within 3 seconds (Shinn & Shinn, 2002b, p. 13).

Overview of the Methodology

This study utilized a quantitative research design. The population for this study was students enrolled in grades 3, 4, and 5 in NKCS. The test data utilized in the study was from the 2010-2011 school year. Data from the MAP and CBM assessments were gathered at the district level.

Multiple regression models were designed for this study to determine the extent to which the independent variables for third, fourth, and fifth grade students predicted communication arts and math MAP assessment scores. Stepwise regression modeling was chosen in order to identify the subset of independent variables to find the best model for predicting the dependent variable (Lunenburg & Irby, 2008). The independent variables for this study were fall, winter, and spring R-CBM, Maze-CBM, and M-CBM scores, gender (male/female), SES (free/reduced and full pay), and race (white/non-white).

Organization of the Study

This study is comprised of five chapters. The introduction chapter provided an overview of the study that included the following sections: problem statement, background of the study, significance, purpose statement, delimitations, assumptions, research questions, definition of terms, and overview of the methodology. The literature review chapter provides a background of research that centers on CBM development history, teacher perceptions of CBM scores, and current research findings for the different CBMs. Chapter three includes the methodology used for this study. The fourth chapter presents the results of the statistical analysis used to address the research questions. Chapter five contains a study summary, discussion of the findings, practical implications for action, recommendations for further research, and a conclusion. The references and appendices follow chapter five.

Chapter Two

Review of the Literature

Over the last four decades, many studies have shown the relevance of Curriculum-Based Measurements (CBM) as an evaluation of student learning. This chapter is dedicated to discussing relevant literature associated with this study of CBM as a formative assessment tool. A comprehensive history of CBM development and changes are presented in the chapter. The CBMs discussed in these sections are the R-CBM, Maze CBM, and the M-CBM. A detailed account of research that utilizes CBM assessment data completes the review of literature chapter.

Uses of Formative Assessments

As described in chapter one, formative assessment data is a method for teachers to utilize when planning for students' instructional needs. This differs from the use of a summative assessment, which students take as an end of course test or end of year exam. Fore, Boon, Lawson, and Martin (2007) clarified the ideas behind summative assessments by stating "Summative evaluation is important as a measure of accountability (i.e., to what degree are students meeting established standards), but does not offer the feedback needed for teachers to make day-to-day adjustments in their teaching" (p. 325). Formative assessments (i.e., CBMs), by contrast, when analyzed within the general education classroom, are designed for evaluation and monitoring of student growth during the instructional process to impact instruction (Deno, 1985). Stiggins et al. (2006) described formative assessments as "assessments *for* learning" which provide "diagnostic/planning information" for teachers, students, and others (parents, etc.) (p. 128).

Prior to the use of CBMs, observation/teacher judgment was the most common form of formative assessment utilized by teachers (Salmon-Cox, 1981). There was a limited amount of commercially licensed assessment material available to teachers at this time (Deno, 1985). Deno (1985) stated that "The fact that teachers rely on informal observations of student performance to evaluate student progress raises the questions of reliability and validity of those judgments" (p. 220). Salmon-Cox (1981) found that when teachers focused their assessments solely on observation, more students were mastering the curriculum than actually had; thus, the teachers were incorrectly assessing student achievement. Therefore, the informal assessment connection between assessment and instructional decisions regarding curriculum were also inaccurate for this group of students (Salmon-Cox, 1981).

In addition to observation, teacher practice included the creation of assessments to determine student mastery of instructional objectives taught (Fuchs, 2004). Fuchs (2004) stated that teachers in the 1970s developed a sequential order for the objectives that students would need to master within a year; additionally, "When a student achieves the mastery criterion for an objective, the teacher simultaneously shifts instruction and assessment to the next skill in the hierarchy" (p. 188). Thus, learning was built around specific objectives; furthermore, when students mastered the material the teacher would move on and not assess this skill again. In contrast, "every CBM test administered within an academic year represents the entire year's curriculum in the same way" (Fuchs & Fuchs, 1991, p. 8).

Formative assessments, when administered and analyzed throughout the school year, provide teachers with a database of information on each student. Fuchs and Fuchs

(1986) conducted a meta-analysis of the effects of systematic formative evaluations. Their results showed an average increase in achievement of .70 standard deviation units when data from formative assessments assisted in the development and monitoring of instructional programs (Fuchs & Fuchs, 1986).

More recently, Hamilton and Shinn (2003) examined 29 teachers from five Pacific Northwest school districts and their ability to estimate students' reading skills in relation to three assessments: the R-CBM, Maze CBM, and the Comprehension Oral Question Answer test. The teacher and student participants were from five rural schools, five suburban schools, and fifteen urban school settings (Hamilton & Shinn, 2003). Teacher participants identified two students from their third grade classrooms. The first student could read fluently, but had comprehension difficulties; the second student was a peer who had no problems with fluency and comprehension skills (Hamilton & Shinn, 2003). The research findings indicated that the teachers significantly overestimated student ability to perform on the assessments (Hamilton & Shinn, 2003). Therefore, the assessments provided reliable numerical evidence of a student's academic standing, which teacher judgment is not able to provide.

Begeny, Codding, Dunn, Eckert, and Kleinmann (2006) examined the accuracy of student CBM assessment data in the areas of reading and mathematics in relation to teacher reported achievement levels. The study included 33 student participants and two teacher participants from a Northeastern suburban school district. The teachers reported judgments of student academic levels (mastery, instructional, and frustration) based on their observations and prior knowledge from six months of classroom instruction (Begeny, Codding, Dunn, Eckert, & Kleinmann, 2006). In the area of mathematics,

teacher assessment predictions and student assessment data correlation coefficients were low overall with a median score of .16 (no *p* value provided) (Begeny et al., 2006). In the area of reading, teacher assessment predictions were higher with a median correlation coefficient of .72 ($p < .01$) with student assessment data (Begeny et al., 2006). These findings indicate teachers were unable to predict the levels of achievement for their students in mathematics, but were able to do so for reading. The assessments provided the teachers with accurate information about their students not biased by background knowledge.

Unlike summative assessments, formative assessments are measurement tools that teachers use to guide instruction through providing immediate feedback to students, identifying student needs, and planning further lessons. Fuchs and Fuchs (1986) found that teachers who employ formative assessments have higher performing students. Moreover, Salmon-Cox (1981), Hamilton and Shinn (2003), and Begeny et al. (2006) indicated that formative assessments provide accurate numerical data over teacher judgment. Formative assessments provide accurate numerical data over teacher judgment, as found by Salmon-Cox (1981), Hamilton and Shinn (2003), and Begeny et al. (2006).

Formative assessment materials can be purchased when a district or individual implements a specific reading series. However, these are used to assess what is written in a particular unit of study that is laid out in the textbook, not for individual state standards. Stiggins et al. (2006) expressed the following statement when thinking of using assessment created by outside sources:

> When using a test developed by someone else—another teacher, a textbook company, or a test publisher—it is crucial to examine it carefully and adjust it as needed for both the learning targets [curriculum] represented and the amount of emphasis each receives prior to deciding to use the test. (p. 129)

Therefore, these formative assessments may not match what is taught in the classroom. Whereas CBMs assess with the learned goal in mind, there is no need for the adjustment that Stiggins et al. (2006) described since the overall learning of the grade level material is the goal.

CBM: A Formative Assessment

Deno (1985) led "an effort to decrease the separation between measurement and instruction—to make data on student achievement more integral to daily teacher decision making—a program of research was undertaken at the University of Minnesota" (p. 221). Not only does CBM data provide teachers with a means to show mastery, they are also able to make reteaching and enrichment decisions through data analyses. In fact, during the initial six-year span of research at the university's Institute for Research on Learning Disabilities, findings showed that teachers were more effective when they utilized [CBM] evaluation methods to lead decision making on student instruction (Deno, 2003; Fuchs, Deno, & Mirkin, 1984). Since initial development, research has validated the assertion that teachers who utilize CBMs over other methods of formative assessments have higher achieving students (Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Steckler & Fuchs, 2000). CBMs provide teachers with a powerful tool for assessing students and assisting with planning for future academic growth.

Fuchs et al. (1984) conducted an 18-week study in New York Public Schools. They monitored teachers who conducted bi-weekly CBM assessments with their students, utilized the data to make classroom decisions, and graphed student results. Students whose teachers used the CBM to assess their reading achieved higher levels than students did from classrooms where other traditional methods of assessment took place; the respective effect sizes for this study were 1.18, .94, and .99 (Fuchs et al., 1984).

Development of R-CBM. Reading was the first Curriculum Based Measurement to be developed. The goal of the R-CBM was to measure growth in the area of text fluency. According to Deno (1985), the idea of writing comprehension questions based on different texts for students to answer was too time consuming and full of "logistical problems" (p. 222). Therefore, the research team looked to create reliable and valid assessments for reading based on cloze, word meaning, and oral reading tasks (Deno, 1985). The creators looked at a variety of formats and duration of assessments to develop a valid assessment measure (Deno, 1992). The R-CBM is the task created in order to assess oral reading tasks. During the early stages of development, teachers chose probe sets for students to read aloud, which met the criteria laid out by Deno (as described in chapter one). Students assessed using the R-CBM read a passage orally for 1 minute while the assessor marks words read correctly and those that are misread (see definition in chapter one).

Deno's research team worked to create valid assessments. One validity study compared the R-CBM to the Comprehensive Test of Basic Skills (CTBS) (Bain & Garlock, 1992). This comparison showed the concurrent validity of the two measurements, or, if the two assessments would have data that correlated to each other

(Lunenburg & Irby, 2008). Data analysis of the mean correlation between the two assessments was .71 with scores ranging from 0.54 to 0.79 (Bain & Garlock, 1992). Therefore, the R-CBM has concurrent validity with the CTBS.

In order to assess students accurately at each grade level, reading passages need to be equally difficult. Ardoin, Suldo, Witt, Aldrich, and McDonald (2005) explained the need for accurate equivalency by stating:

> If the difficulty in the reading assessments is not equivalent, then the passage will either over- or underestimate the students' "true" performance. For example, if passages given to a student during week two are easier than passages administered in weeks four and five, the students' progress will appear to decline. (p. 2)

However, grade level textbook passages are not the best option. Textbooks are written for a wide range of readers, even when presented as a grade level materials. Readability formulas provide a resource for determining the grade level of reading passages. The majority of readability formula models input syllables, word count, and sentence length. Ardoin et al. (2005) utilized readability formulas to determine that passages in third grade textbooks may have readability ranges from 4.8 to 5.7 (i.e. fourth grade, eighth month to fifth grade, seventh month). Similarly, fourth grade textbooks were found to have ranges from 7.5 to 8.8 (Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005).

With standardization in mind, districts have begun to use the commercially prepared probes from Edformation, Inc. (housed in AIMSweb): "[d]ifferences between probe sets were minor and not systematic" (Christ & Silberglitt, 2007, p. 135). The degree of difficulty of the passages given at each grade level was equated using Lexile scores and student performance data (Howe & Shinn, 2002).

Researchers have found other compelling evidence for the use of the R-CBM in grade 3 versus higher grades. Kranzler, Miller, and Jordan (1999) examined the R-CBM in relation to the California Achievement Test, and the occurrence of gender and racial/ethnic bias of African American and Caucasian students in grades 2-5 (n = 326). Their findings showed no biases at grades 2 and 3, but found significant intercept biases based on race/ethnicity in grades 4 and 5 (Kranzler, Miller, & Jordan, 1999). In addition, significant intercept and slope biases indicated gender bias in grade 5 (Kranzler et al., 1999). Therefore, although the R-CBM is a reliable assessment, there is data to confirm that the R-CBM is biased towards different gender and race/ethnicity groups at the upper elementary level.

Development of Maze CBM. An alternate formative assessment to the R-CBM is the Maze CBM (Maze). The Maze measures student growth in the areas of decoding and comprehension. It requires students to silently read a passage and circle one of three word choices in order to fill in a blank within the passage (see definition in chapter one).

The Maze CBM has changed over time. Initially, the Maze task was untimed and the students would read a passage and make their word choices (Wiley & Deno, 2005). Current Maze CBM studies have "established that timed maze scores are less negatively skewed and likely to increase validity coefficients" (Wiley & Deno, 2005, p. 208). Therefore, current practice is to use a timed assessment to increase the validity of the assessment. However, the question remained: how much time should students be given to perform the assessment?

Jenkins and Jewell (1993) studied a possible relationship of the 60-second Maze assessment to other assessments. Pacific Northwestern students (n = 335) from grades 2-

6 took part in the study that validated a timed Maze assessment with correlations ranging from .66 to .76 with the Metropolitan Achievement Test and correlations ranging from .65 to .76 with the Gates-MacGinitie Reading Tests (Jenkins & Jewell, 1993). Therefore, the Maze, when compared to the Metropolitan Achievement Test and the Gates-MacGinitie Reading Tests, is able to evaluate the reading skills it was developed to assess.

Research of the Maze CBM has established it to be a reliable assessment; however, Brown-Chidsey, Johnson, and Fernstrom (2005) extended the research by examining the relationship between the type of reading passage that a student is given and their scores. Twenty-one Caucasian fifth grade students from a northeastern suburban school district participated in this study by completing a fall, winter, and spring Maze CBM (Brown-Chidsey, Johnson, & Fernstrom, 2005). At each assessment session, the students read from both a controlled passage and a literature passage, resulting in two separate Maze scores (Brown-Chidsey et al., 2005). The scored controlled fifth grade passages, gathered from the AIMSweb program, were compared to the literature passages, which originated from the American Library Association's list of fifth grade recommended reading selections (Brown-Chidsey et al., 2005). Their findings were similar to those of Ardoin et al. (2005) who studied the relation of controlled R-CBM passages to textbook reading passages; the controlled passage had consistently higher scores than the literature passages (Brown-Chidsey et al., 2005). The total possible points varied from passage to passage based on the length of the first sentence; however, each passage was around 250 words in length. For example, the median spring controlled passage received a score of 20.07; whereas, the median spring literature based passage

score was 13.14 (Brown-Chidsey et al., 2005). Therefore, the controlled passages consistently produce higher scores than those chosen from authentic literature.

When creating a Maze measure, the passage source is not the only characteristic that causes concern. After passage selection, it must be prepared for the assessment. The first sentence of the passage remains complete; thereafter, every seventh word is omitted. Three word choices are placed in the text for the student to choose from the correct replacement and two distractors. Parker and Hasbrouck (1992) described the differences between Maze studies are "in subjects, materials, and procedures [that] do not allow comparison on the effects of different types of distractors" (p. 201). However, an item analysis completed by McKenna and Miller (1980) showed item difficulty was not affected by word length and visual configuration when third through sixth grade students were assessed with the Maze.

Fuchs and Fuchs (1992) compared cloze techniques and retell methods to the Maze task. Their goal was to show which of the three assessments would provide teachers with a quick and reliable alternative to the R-CBM. The participants in this study were 63 special education students with below grade level reading skills. Fuchs and Fuchs (1992) set the following guidelines for Maze distractors:

> Distractors had to be the same length, one letter shorter, or one letter longer than the correct replacement. Distractors could not (a) make contextual sense, (b) rhyme with the correct replacement, (c) be close in either sound or letter configuration to the correct replacement, (d) be a nonsense word, (e) require the student to read more than 1.5 lines ahead in the passage to eliminate, or (f) be so

high in vocabulary that the student might mistake the distractor for a nonsense word. Each passage was edited twice for compliance with these criteria. (p. 55)

Their results showed that the Maze was a more adequate method for assessing student-reading growth over the cloze and retell methods. When using these guidelines for the creation of Maze CBM, criterion validity is present (Fuchs & Fuchs, 1992).

Within the reading process, syntax and semantics are important language cueing systems. Deno et al. (2002) responded to this concern by studying the placement of the omitted word. Students were administered alternate forms of the assessment with the omitted words at the end of each sentence. Results indicated no significant differences between student scores, suggesting that the placement of the omitted word does not affect student performance on the assessment (Deno et al., 2002).

One advantage of the Maze over R-CBM is that it does not require that a professional administer the assessment to students individually. Instead, students are able to complete the evaluation in groups or via computerized assessments. As a result, the Maze CBM is less time consuming for teachers to administer.

When Fuchs and Fuchs (1992) conducted their study comparing the Maze against other assessment measures (see previous discussion), they also researched teacher and student perceptions. The teachers questioned were very satisfied with the Maze CBM because, "the measure seemed to reflect the multiple dimensions of reading, requiring pupils to demonstrate decoding, comprehension, and fluency" (Fuchs & Fuchs, 1992, p. 50). In addition, the 63 student participants also rated the Maze CBM as an assessment that they liked taking with a mean of 1.69 on a 3-point scale, SD = .75 (Fuchs & Fuchs, 1992, p. 55).

The Maze CBMs are assessments that can measure reading comprehension. One major difference between the Maze CBM and the R-CBM is whether the child reads the material silently or aloud. In 2011, Hale et al. questioned if comprehension, as measured by the Maze CBM, was affected by how the child reads the material: orally or silently. Eighty-nine first and second grade students from the Southeastern United States participated in the study (Hale et al., 2011). The authors stressed several limitations to their work. These included that there was no way to determine if a student was choosing not to read when students are allowed to read silently and they felt that their sample size was relatively small with a lack of student diversity (Hale et al., 2011). Notwithstanding, Hale et al. (2011) concluded "the current study suggests that requiring students to read out loud, as opposed to silently, does not hinder student comprehension" (pp. 10-11). For practitioners, this information is important to show that scores are not affected by the manner of reading that the child chooses.

The final step in administering the assessment is the scoring. Pierce, McMaster, and Deno (2010) examined whether different scoring procedures affected the Maze CBM score. They collected data from 199 students from two Midwestern school districts in the fall and spring of the same school year (Pierce et al., 2010). The participants in the study were "poor readers…, at risk, or identified as having disabilities in reading" (Pierce et al., 2010, p. 153). This focus was chosen since at-risk students are monitored more often than general education students with CBMs and thus, if scoring variance affects reliability, then students are affected (Pierce et al., 2010). With frequent formative assessments, growth trends or dips are apparent and allow for quick instructional changes for at risk student populations. Pierce et al. (2010) explained that "all probes [reading

passages] were scored using five different methods (one that involved simply counting all correct Maze choices and four that were meant to reduce the effect of random guessing)" (p. 154). The results of the study showed that any of the five scoring procedures used with the participants would yield highly correlated and valid scores for the students (Pierce et al., 2010). Since the monitored assessments were performed only twice throughout the year, the researchers were unable to show the slope of growth for the participants based on the scoring procedures; however, the scores were found to be technically sound (Pierce et al., 2010). The findings of the study showed that standardization of Maze CBM occurs not only when passages are marked for only correct scores, but also when monitoring for random guessing.

Development of M-CBM. The M-CBM went through a similar process of development as the R-CBM and Maze: format development, duration of the assessment, and a process to define what the assessment was to measure. Clarke and Shinn (2004) notes therefore, the "M-CBM, like other CBM measures, is based on a validated, standard, simple to administer and score, short duration fluency measure where students write the answers to computational problems" (p. 237).

Within mathematics, there are two broad constructs commonly referred to within educational literature: computational problems and application based problems. Computational problems require students to know how to work with mathematical concepts, strategies, and facts (Howell, Fox, & Morehead, 1993). In contrast, application problems use the mathematical strategies, concepts, and facts to solve a given problem. According to Thurber, Shinn, and Smolkowski (2002), M-CBMs were "designed to serve as a measure of general math achievement, not specifically as a measure of only

computation or applications" (p. 500). Therefore, the assessment's design is not to just assess computation or application, but a mixture of both.

However, Thurber et al. (2002) conducted a study "to examine the relation of the M-CBM to the constructs of general mathematics achievement, computation, and application from a theoretical perspective using confirmatory factor analysis" (p. 500). Their study included 207 Northwest fourth grade students from four elementary schools with an almost equal male to female ratio; general education students made up 74% of the population (Thurber et al., 2002). The findings of the study confirmed that the M-CBM is a computational, rather than an application, assessment, although with a lower than expected evidence of validity (Thurber et al., 2002). Nonetheless, Rutherford-Becker and Vanderwood (2009) maintain "there is a degree of dependence among the math constructs, applied and computational, indicating that the skills in one area are necessary for success in the other" (p. 24).

Further work of Fuchs (2004) described that with the development of mathematical tasks, two broad approaches are used. The approach of curriculum sampling applies to the development of computational methods (Foegen, Jiban, & Deno, 2007). This method takes an overall look at the curriculum for a particular grade when developing the assessment. For example, in second grade, "a larger proportion of addition and subtraction problems and, at sixth grade, a sampling that includes more advanced skills are used, such as division of decimals or addition of fractions" (Foegen et al., 2007, p. 121). This type of assessment provides teachers with measurements that tie directly to the curriculum taught. However, the assessment is geared to one grade level's

curriculum; therefore, having multitudes of assessments to administer maintains the formative nature of the assessment.

The M-CBM is a standardized assessment, which measures the growth of computational mathematical ability. For the M-CBM, the correct answers, as described by Christ et al. (2005), are "quantified as the number of digits written correctly per unit of time… Digits correct are defined as any digit written by the target student that is in the correct place value within the solution of a mathematical problem" (p. 615). Research performed by Christ et al., (2005) examined what length of time should be utilized in order to provide valid results for this assessment. Fourth and fifth grade general education students (n = 104, equally divided) from the southeastern United States participated in the study. The findings of this study indicated that for general screening measures (assessing for grade level norms) a 1-minute assessment is reliable but for higher stakes assessment needs (assessing for individual diagnosis), a 4-minute assessment is needed (Christ, Johnson-Gros, & Hintz, 2005).

Since "mathematics is generally accepted as more skill specific than reading, content for CBM mathematics tests is derived by determining the grade-level skills deemed important in the student's curriculum" (Steckler, Fuchs, & Fuchs, 2005, p. 798). However, Clarke and Shinn (2004) specified that the M-CBM requires students to have a prior knowledge of mathematics. Until students reach this point in their education (mid to late first grade), their initial scores will remain zero (Clarke & Shinn, 2004). After this prerequisite, the assessments are then able to provide teachers with valuable information on which to base their instruction.

The nature of the CBM administration procedures calls for students to take the assessments throughout the year to assess academic growth. Christ and Schanding (2007) researched a possible limitation of the assessment: familiarity with procedures. Ninety second through fifth grade students, from the southeastern region of the United States, participated in the study (Christ & Schanding, 2007). The purpose of the study was to determine if student scores would be influenced by novelty (a new assessment event – no prior exposure to the M-CBM), neutral (taking the M-CBM with previous exposure), or reward conditions (promise of a Popsicle if the scores increased) (Christ & Schanding, 2007). A "reduced level of student performance [was] observed during the novel condition as compared to performance during neutral or reward conditions. These results suggest that students accurately completed computational tasks with less fluency when the assessment conditions were unfamiliar" (Christ & Schanding, 2007, p. 153). These findings are important for M-CBM data analysis. If students are new to the assessment, their first exposure and administration scores will be lower than students who have had a prior experience with the assessment (Christ & Schanding, 2007). This is necessary for rank ordering of current students and new students moving into a school. Finally, it is interesting to note that the scores for students during the neutral and reward conditions were similar. Christ and Schanding (2007) stated "[t]his implies that most students do their best to perform well during M-CBM administrations" with or without a reward (p. 156).

Fuchs, Fuchs, Hamlett, and Stecker (1991) found that when teachers utilized M-CBM to formulate instruction, their students had higher outcomes, over time, when provided with support/consultation. Within the study, placement of 33 teachers occurred

into one of three groups: a control group, which did not utilize CBM data, a treatment group with no support for student administered CBM assessment data, and a treatment group with support for utilization of the administered CBM assessment data (Fuchs, Fuchs, Hamlett, & Stecker, 1991). According to researchers, both treatment groups that administered the CBM within their classrooms utilized the data for improving instruction; however, the CBM group with support had superior student achievement over the other two groups (Fuchs et al., 1991). Therefore, support and consultation assistance for teachers is vital for student success when M-CBM formative assessments are a part of classroom practices.

Growth Rate Research

CBMs provide teachers with a formative assessment, which is a reliable and valid indicator of a student's skills (National Center for Response to Intervention, 2010). However, students' summative growth achievements are monitored with the state level assessments. Therefore, it is imperative that the CBM provides teachers with assessment data that they can use to guide instruction and details on how students will perform on the state assessments. Graff, Jenkins, and Miglioretti (2009) stated in their article *Estimating Reading Growth Using Intermittent CBM Progress Monitoring* that "Achievement tests are insensitive as measures of short-term growth and are time-consuming for students to complete – precisely the reasons Deno devised CBM in the first place" (p. 153).

The R-CBM measures oral reading fluency by measuring how many words a student read correctly within a minute. Therefore, in order to show that they have grown in fluency, the number of words read correctly would increase with instruction. This

increase is illustrated as a slope of improvement over time (the greater the slope of improvement, the greater the amount of student progress).

A goal of monitoring growth rates is to determine if students are learning within their present environment or if they need a change. Graff et al. (2009) researched R-CBM growth rates for this reason. Their study monitored 41 third through fifth grade students from eight schools within the Seattle area over a 9-week time span (Graff, Jenkins, & Miglioretti, 2009). Results of their data analysis showed that the first, third, and last weeks' data generated growth scores that aligned with the researchers' belief of the true slope. The data "pass[ed] one test of validity in that they generate growth scores that are comparable in size to our best estimate of true grow…[and] we were able to create a strong validity criterion of reading growth" (Graff et al., 2009, pp. 159, 161).

Christ, Silberglitt, Yeo, and Cormier (2010) agreed that there is validity to justify the analysis of R-CBM growth rates throughout the academic year. However, the analysis of R-CBM growth rate data shows a seasonal effect that determines when the most growth takes place within a calendar year (Christ, Silberglitt, Yeo, & Cormier, 2010). Their sample included 4,824 second through sixth grade students from seven elementary schools, in five rural and suburban Midwest school districts (Christ et al., 2010). The seasonal change for general education students is shown in Table 2. The fall season refers to the time between the fall and winter CBM assessments and the spring season refers to the time between the winter and spring assessment. The fall season scores consistently showed more growth in words read per minute than the spring season.

Table 2

Descriptive Statistics for Seasonal Growth Change

Grade Level	Fall Season *M (SD)*	Spring Season *M (SD)*
Second	1.71 (.73)	1.02 (.64)
Third	1.38 (.79)	0.97 (.75)
Fourth	1.19 (.72)	0.78 (.70)
Fifth	1.12 (.76)	0.83 (.75)
Sixth	0.89 (.71)	0.74 (.75)
Total	1.26 (.74)	0.87 (.72)

Note. Adapted from "Curriculum-Based Measurement of Oral Reading: An Evaluation of Growth Rates and Seasonal Effects among Students Served in General and Special Education," by T. J. Christ, B. Silberglitt, S. Yeo, & D. Cormier, 2010, p. 453.

Christ et al. (2010) showed statistical data for R-CBM growth rates both within and across grade levels. Their analyses indicated that the amount of growth slowed as students progressed from primary grade levels. Christ et al. (2010) stated "Both visual analysis and statistical analysis of growth models support the conclusion that there…was generally robust [growth] in the lower grades" (p. 457). However, as students reach fifth grade, the R-CBM has not shown to assess student reading with as much validity as other measures (Brown-Chidsey et al., 2005).

Shin, Deno, and Epsin (2000) studied the relationship between student growth rate and the Maze CBM. Their study focused on the reading growth rate of second grade students. They assessed each student once a month for nine months to discover the

technical adequacy of the Maze CBM and growth rates and found "The correlation between monthly Maze scores with 1- to 9-month intervals ranged from .69 to .91 with a mean of .81" (Shin, Deno, & Epsin, 2000, p. 167). Thus, according to the results, the Maze is a reliable source of data for student growth over time.

Shin and Lee (2007) examined the Maze CBM growth rates of 273 urban, Midwest general education students in order to show how schools, districts, and states agencies could use a combination of Maze CBM data and hierarchical linear modeling to create growth rate norms. Shin and Lee (2007) reported that the linear growth rate (π_{1i}) was .87 in the study. This is moderately strong evidence for the reliability of the instrument. Upon completion of the data analysis, the mean growth rates "in grades two to six were [respectively] .45, .68, .62, .60, and .33 increases per month on the Maze [CBM] task" (Shin & Lee, 2007, p. 360). Shin and Lee (2007) considered these growth rates as minimum standards for utilization in urban school settings. However, if applied, the growth rates are a means for goal setting and progress monitoring of student success (Shin & Lee, 2007).

Along with the Maze and the R-CBM, it is also important to look at the growth rate research centered on the M-CBM. This research focused on whether the CBMs were a valid approach for showing student achievement over time. One study focused on the validity of all three measurements. Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) performed a 2-year study to determine if growth rate results at both weekly and monthly intervals were valid. During the first year of their research, they studied M-CBM and R-CBM assessment results to calculate the validity of utilizing the assessment weekly for students in grades 1-6 from the upper Midwest (Fuchs, Fuchs, Hamlett, Walz, &

Germann, 1993). There were 177 students assessed utilizing the M-CBM and 117 students assessed utilizing the R-CBM. Within the study, students assessed using the R-CBM once a week over the span of an academic school year showed, on average, positive linear growth. The estimated gains for words read correctly each week were: 1.5 in first and second grade, 1.0 in third, 0.85 in fourth, 0.5 in fifth, and 0.3 in sixth (Fuchs et al., 1993). Interestingly, although there was growth indicated in each grade level, there was a negative trend of words gained per week across the grades (Fuchs et al., 1993). The M-CBM assessment showed a weekly growth rate increase which ranged from .2 in grade 2 to .77 in grade 4 (Fuchs et al., 1993).

Within the second year of the study, the researchers employed the M-CBM and the Maze assessment to determine if growth rates for the assessments were valid if completed monthly. Although they assessed the 257 students with the Maze and 1,208 students with the M-CBM monthly, the findings of the study were reported as weekly rates of improvement (Fuchs et al., 1993). Unlike the R-CBM, the Maze data slopes did not vary significantly due to grade level and found a mean growth rate of .39 words (Fuchs et al., 1993). Additionally, the M-CBM assessment showed a weekly growth rate increase that ranged from .28 in grade 2 to .74 in grade 5 (Fuchs et al., 1993). Thus, the overall findings of the study showed that the three separate assessments all indicated valid growth rates for student achievement when utilized throughout the school year.

CBM Research Showing Predictability of High Stakes Testing/State Assessments

Research has concluded that CBMs are statistically reliable and valid forms of measuring student achievement (National Center for Response to Intervention, 2010). Researchers have now moved towards finding further information regarding the uses of

the CBM. Within general education classrooms, teachers review CBM data to monitor students who are at risk of not mastering the grade level curriculum and, in turn, failing state assessments. Research conducted in this area show varying results for using R-CBM and Maze scores for predicting state assessment scores. Less research has been conducted in the area of predicting mathematical state assessment scores with the M-CBM. One important by-product of this research is information concerning the possible demographic bias of the CBM assessments. Hintze, Callahan, Matthews, Williams, and Tobin (2002) explained that this occurrence existed because "[t]esting for such bias involves the use of regression analysis as a means of predicting performance on a secondary measure (e.g., test of reading comprehension) on the basis of a person's CBM score" (p. 540). Not all studies researching CBM and state assessments monitored for CBM bias; however, if researchers reported this information, it is included.

Stage and Jacobsen (2001) studied whether the R-CBM predicted the performance of 173 fourth grade students on the Washington Assessment of Student Learning (WASL). The WASL had a predicted base failure rate of 20% and base passage rate of 80%; however, through the data analysis of R-CBM scores in the fall, winter, and spring, these researchers were able to increase predictive power by 30% (Stage & Jacobsen, 2001). The researchers utilized the CBM scores to monitor student achievement through the fall, winter, and spring. When Stage and Jacobsen (2001) compared the CBM scores and the WASL assessment scores, instead of the blanket base prediction of 80/20 pass-fail rate, they predicted the student performance on the WASL with 30% more accuracy than the base rates.

McGlinchey and Hixson (2004) replicated the study "of Stage and Jacobsen (2001), with a different state fourth grade reading test across 8 years, a much larger sample of students, and a more diverse student population" (pp. 194-195). Their work monitored 1,362 fourth grade participants in a Michigan urban community from the years 1994-2001. Additionally, McGlinchey and Hixson's (2004) study included, "across the district, the non-Caucasian population was 52%, and the 'free and reduced' lunch status (an indicator of socioeconomic need) was 60%" (p. 195). The majority of the study took place within one school; however, year 4 of the study monitored fourth grade data across the whole school district. The construction of the study included the R-CBM's words per minute as an independent variable to determine if students reading more than 100 words per minute would receive a satisfactory/pass score on the Michigan Educational Assessment Program (MEAP) (McGlinchey & Hixson, 2004). McGlinchey and Hixson (2004) determined that with the cut score at 100 words per minute, "the specificity of the cut score for identifying students who did achieve satisfactory scores was 74%. The sensitivity of the cut score for identifying students who did not achieve satisfactory scores was 75%" (p. 198). With the longevity and methods of the study, they added to the research focused on the cut scores and their predictive validity of the R-CBM and state assessments. Although the authors did not evaluate the data by demographics, they stated that the CBM assessments could be a resource for working with students from "districts with a high percentage of children from low-income backgrounds as in the current study" (McGlinchey & Hixson, 2004, p. 202).

In partial fulfilment of graduate level work at the University of Florida, Ax (2004) summarized findings of nine R-CBM correlations with state tests to be statistically

significant and highly correlated. Table 3 provides the cut scores and accuracy in predicting a passing score on the individual state assessments. If the students received or surpassed the cut score provided on the R-CBM in Table 3, then they received a passing score on the state assessment.

Table 3

Cut Scores for Predicting Passing Score in Grade 3 on State Achievement Tests

State	Cut Scores	Accuracy in Predicting Pass
Colorado	100	90%
Oregon	110	99%
Illinois	110	99%
Michigan	100	72%
Pennsylvania	114	93%
North Carolina	110	100%
Washington	100	90%
Florida	110	91%
Ohio	110	Unavailable

Note. Adapted from "Relationship Between Oral Reading Rate and Statewide Achievement Test Mastery for Third Grade Students," by E. E. Ax, 2004, p. 22.

The data show that of the states presented, the R-CBM's ability to present a passing score on the North Carolina test was 100%; this was the highest rate shown (Ax, 2004). In contrast, Michigan's assessment pass rate was the lowest at 72%.

Hintze and Silberglitt (2005) researched the predictive validity of the R-CBM to the Minnesota Comprehensive Assessment (MCA) at the end of third grade. The

participants in this study included 1,766 elementary school students who were administered the R-CBM eight times between the fall of grade 1 and the spring of grade 3. According to the results of their data analysis, the R-CBM "appears to be an efficient method for predicting performance on high-stakes tests demonstrating the ability to predict those students who are likely to pass reading portions of such tests as far back as first grade" (Hintz & Silberglitt, 2005, p. 382). Districts utilizing the R-CBM are then able to monitor student achievement levels in order to estimate how their students will perform on high stakes tests. A secondary finding of their research indicated that the R-CBM cut scores were a more accurate measure of student benchmarks within and between grade levels than the MCA (Hintz & Silberglitt, 2005).

Silberglitt, Burns, Madyun, and Lail (2006) reviewed the relationships between the Minnesota state assessments and the R-CBM and Maze assessments. Participants included 5,472 students from five Minnesota school districts in grades 3, 5, 7, and 8 (Silberglitt, Burns, Madyun, & Lail, 2006). One limitation of the study was 94% of the participants were Caucasian (Silberglitt et al., 2006). The study utilized two Minnesota state mandated assessments. Students in grades 3, 5, and 7 took the Minnesota Comprehensive Assessments-Reading (MCA-R) and grade 8 students took the Basic Standards Test-Reading (BST-R). Eighth grade students were able to take the BST-R more than one time; however, only the scores from the first administration were utilized (Silberglitt et al., 2006). Statistically significant ($p < .001$) relationships were found between the MCA-R, the BST-R, and the R-CBM; in addition, all coefficients were equal to or greater than .50 (Silberglitt et al., 2006). Next, these researchers investigated if a relationship existed between grade level R-CBM scores and the MCA-R assessment for

grades 3, 5, 7, and 8 (Silberglitt et al., 2006). They found that with R-CBM the relationship between coefficient scores were significant in third grade, but decreased in significance as the grade level increased (Silberglitt et al., 2006). This finding has important implications for educators since it can be assumed that the relationship between the R-CBM and state assessments continues as students advance through grade levels (Silberglitt et al., 2006). Their final question "addressed significant differences in relationship magnitude among… reading assessments, fluency [R-CBM] and Maze, as they compared to state accountability test scores" (Silberglitt et al., 2006, p. 532). The Maze assessment was administered to 7th and 8th grade students in the study and findings indicated statistically significant and moderate coefficients between the Maze and MCA-R scores for these students at .54 and .49, respectively (Silberglitt et al., 2006). The same analysis for the R-CBM and the MCA-R produced results of .60 for 7th grade students and .51 for 8th grade students (Silberglitt et al., 2006). Researchers "found no significant difference between coefficients associated with R-CBM fluency and Maze for either grade" (Silberglitt et al., 2006, p. 532). Therefore, both the Maze and R-CBM scores could be utilized as comparisons to the MCA-R.

School stakeholders have an interest in how students are developing their reading skills prior to the mandated tests, which occur in third grade. In order to provide this information, districts monitor student development with different formative assessments, such as the R-CBM. Hintz and Silberglitt (2005) stated that R-CBM scores from first grade students could predict a passing score on standardized assessments.

Wanzek et al. (2010) developed a study to determine if student R-CBM scores across first, second, and third grades predicted mastery on the Texas Assessment of

Knowledge and Skills (TAKS) and the Stanford Achievement Test, 10th edition (SAT-10) for third grade students. This study included 461 participants from one Texas school district (six elementary schools) with a high percentage of free and reduced lunch (75%) and minority students (66% Hispanic, 19% African American, 2% other); however, if students were in the district's bilingual classroom, they were excluded from the study (Wanzek et al., 2010). Students were not included if they did not take an R-CBM assessment prior to the spring semester of third grade, the TAKS, and the SAT-10 (Wanzek et al., 2010). These choices ensured that the student population included only students who participated in the necessary assessments and excluded the variable of a language barrier. Of these students, "87.0% achieved a passing score on the TAKS (scaled score of 2100 or higher) and 41.5% scored at a proficient level or higher on the SAT-10 (scaled score of 634 or higher)" in third grade (Wanzek et al., 2010, p. 72). The researchers determined that the R-CBM was a predictor for student success on the TAKS and SAT-10; however, the R-CBM cut score to show this prediction varied for two separate assessments (Wanzek et al., 2010). Assessment procedures occurred in the winter and spring of the sample's first grade year and three times a year thereafter for their second and third grade years. They found that students needed to increase by "6.7 [correct] words per minute between each measurement period in first through third grades to remain on track and have a probability of .8 of proficiency on the TAKS" (Wanzek et al., 2010, p. 74). However, to have the same proficiency on the SAT-10, students needed to increase their words per minute by 16.4 in between each R-CBM probe (Wanzek et al., 2010, p. 74). The R-CBM, as a formative assessment, provides teachers with information to assist them with designing instruction to assist with this increase. Also, all students

who were proficient on the SAT-10 passed the TAKS (Wanzek et al., 2010). The researchers did not single out their sample demographic statistics; however, it is important to note that free and reduced lunch and minority participants showed success in the standardized assessments (Wanzek et al., 2010).

State assessment measures are important throughout the United States to monitor adequate yearly progress towards federally mandated No Child Left Behind. In California, Kranzler et al. (1999) began an examination of racial, ethnic, and gender bias to determine if the R-CBM was biased against African American and/or Caucasian students. Their research evaluated the R-CBM and the California Achievement Test (CAT) scores of 225 Caucasian and 79 African American students (n = 326, 170 boys and 156 girls) in grades 2-5 (Kranzler et al., 1999). Overall, they found that "CBM is a strong predictor of reading comprehension at each grade level" on the CAT (Kranzler et al., 1999). However, they did find evidence of bias within their multiple regression analyses. Gender data showed no significant differences between scores on the R-CBM and the CAT reading comprehension assessment, except in grade 4 where girls received higher mean scores on both measures than did boys (Kranzler et al., 1999). Their racial/ethnic data analysis determined that no bias was present within grades 2 and 3; in contrast, the CBM predicted that African American students in grades 4 and 5 would receive higher scores than they actually achieved (Kranzler et al., 1999). Therefore, they underestimated the scores of the Caucasian students in these grades. These findings were shown through evidence of both intercept and slope bias; therefore, the findings of Kranzler et al. (1999) show that within certain situations, the possibility of R-CBM bias exists.

Hintze et al. (2002) followed up on the research of Kranzler et al. (1999) with different findings. The participants in their study included 136 urban Northeastern students in grades 2-5 (66 male, 70 female, 65 African American, and 71 Caucasian), also accounting for the socio-economic status (SES) and age of the students (Hintze et al., 2002). The assessments utilized by the researchers included the R-CBM and the Woodcock Johnson Psychoeducational Battery-Revised (WJ-R), which is a comprehensive standardized assessment of both cognitive and academic abilities (Hintze et al., 2002). Hintze et al. (2002) determined the following through their data analyses:

> Results suggested that the CBM oral reading fluency metric was not biased with respect to ethnicity or SES. That is, once developmental level (i.e., age), oral reading fluency, and SES were accounted for, ethnicity predicted very little with respect to reading comprehension abilities. Moreover, the data were consistent with the extant empirical literature indicating that CBM in reading is a strong indicator of generalized reading performance including reading comprehension. (p. 545)

Therefore, Hintze et al. (2002) and Kranzler et al. (1999) differ in their findings that the R-CBM is biased concerning gender and racial/ethnic background. However, both agree that the R-CBM is a predictor for high stakes achievement scores.

Merino and Beckman (2010) also researched the R-CBM assessments in relation to a high stakes assessment; however, they developed their study to include the Maze. Their research focused on these measures as predictors for academic success on standardized assessments in Nebraska. Their research questions asked if a spring administered R-CBM, Maze, or R-CBM and Maze combination could predict the

Measures of Academic Progress scores (Merino & Beckman, 2010). The Measures of Academic Progress assessment, as defined by Merino & Beckman (2010), "is a high stakes assessment published by the Northwest Evaluation Association… [to assess] word recognition and decoding, literal comprehension, interpretive comprehension, and evaluative comprehension" (p. 90). Their participants included 376 elementary school students in grades 2-5 from diverse backgrounds. Multiple regression analyses showed that the R-CBM and the combination of R-CBM and Maze assessments had statistically significant predictions of student Measures of Academic Progress scores in each grade assessed (Merino & Beckman, 2010). However, the Maze assessment alone did not show statistically significant predictions for the Measures of Academic Progress in any of the grades assessed (Merino & Beckman, 2010). With these results in mind, Merino and Beckman (2010) questioned the need to administer the Maze assessment as a predictor for students who take the Measures of Academic Progress as their state's high stakes assessment. The R-CBM assessment alone produced scores that were a better predictor of the Measures of Academic Performance (Merino & Beckman, 2010). Therefore, other school populations with similar demographics are able to consider this when deciding what assessment to utilize.

Research showing the predictive validity of CBM and achievement on high stakes testing has predominantly focused in the area of literacy. However, some researchers have conducted studies that focused on both curriculum areas and state assessments. Shapiro, Keller, Lutz, Santoro, and Hintze (2006) studied whether M-CBM and R-CBM were predictors of student outcomes on the following standardized assessments: Pennsylvania System of School Assessment (PSSA); Stanford Achievement Test, 9th

edition (SAT-9); Metropolitan Achievement Test, 8th edition (MAT-8); and Stanford Diagnostic Reading Test (SDRT). The participants in the study included students from two districts in Pennsylvania and excluded students with active Individual Education Plans (except for Gifted and Talented services) (Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). Student participants from District 1 consisted of 617 (for reading assessments) and 475 (for math assessments) 3-5 grade students from 14 elementary buildings (Shapiro et al., 2006). In District 1, all participants took the CBM in reading/math, students in grades 3 and 5 took the PSSA, students in grade 4 took the MAT-8 reading, and students in grade 5 took the SDRT. District 2 consisted of 431 participants from grades 3-5 from eight elementary school buildings, for both math and reading assessments (Shapiro et al., 2006). Within District 2, all participants took the CBM in reading/math, students in grades 3-5 took the PSSA, and students in grade 4 took the SAT-9 in reading and math (Shapiro et al., 2006). Collection of CBM data occurred in the fall, winter, and spring. In relation to the PSSA, all M-CBM scores were statistically significant except for the fall score in District 1; however, fall scores were consistently lower (Shapiro et al., 2006). Shapiro et al.'s (2006) results showed "good support for the use of CBM computation as a moderate predictor of outcomes on state assessments… [and were] consistent across winter and spring assessment periods, grades, as well as across two school districts" (p. 31). PSSA and R-CBM scores had statistically significant correlations, consistently in the .50 range (Shapiro et al., 2006). With regard to the other standardized assessments administered (MAT-8, SAT-9, and SDRT), relationships were moderate to strong. Correlations were in the .70s in reading between R-CBM and both the MAT-8 and SAT-9 (Shapiro et al., 2006). Statistical analysis for the spring M-CBM and the SAT-9 found

a strong relationship at .688 ($p < .001$) (Shapiro et al., 2006, p. 30). Results from the study determined that M-CBM and R-CBM had strong correlations to a state assessment and to several norm-referenced standardized assessments across the two districts.

Keller-Margulis et al. (2008) conducted a subsequent study to examine the "long-term diagnostic accuracy of CBM in reading and mathematics" within a Pennsylvania school district (p. 374). The school district administered the PSSA to students in grades 3-5; in addition, they administered the *TerraNova*, Second Edition to students in grade 4 (Keller-Margulis et al., 2008). The purpose of their study was to determine if a one- to two-year long-term relationship existed between student CBM (fall, winter, and spring), PSSA, and *TerraNova* scores in both reading and mathematics. Within the area of reading, the R-CBM was the assessment measure. The "[s]lope and statewide achievement test 1 year later did not demonstrate a strong relation in reading. However, the 2-year relation indicated that growth in the early grades (first and second grade) was related to future success" on the PSSA and the *TerraNova* (Keller-Margulis et al., 2008, p. 385). Overall, the findings showed a low predictive power for the fall and winter CBM. However, Keller-Margulis et al. (2008) found that the "pattern of results with reading CBM indicated that across 1 and 2 years, the single point [spring] CBM data were moderately and positively correlated with the statewide achievement test as well as nationally known, norm-referenced achievement measure" (p. 385). The second grade spring reading CBM predicted the *TerraNova* assessment in fourth grade with 87% accuracy (Keller-Margulis et al., 2008, p. 383). Keller-Margulis et al. (2008) cautioned that "inconsistent results for the math CBM slope related to the PSSA and *TerraNova* indicated that the math CBM may have behaved in a manner distinctly different from the

reading CBM" (p. 386). The second grade spring mathematics CBM predicted the *TerraNova* assessment in fourth grade with 77% accuracy (Keller-Margulis et al., 2008, p. 384). However, overall, the M-CBM assessments did not produce positive predictive power over 62% on any of the assessments. The MAP assessment used for the current study includes a portion of the *TerraNova.*

In addition to the preceding studies, Shapiro (2012) further added to the research correlating the M-CBM and the Pennsylvania System of School Assessment (PSSA) by researching the relationship between these two assessments with a computer-adaptive test (CAT). The CAT used within Shapiro's (2012) study was STAR-Math, which is "based on the accuracy of student responding and examine the key skills that lead to effective mastery of competencies within an academic domain" (p. 296). Therefore, if a student responds correctly to an item, the computer will generate an item of greater difficulty within the same skill set, but if the student answers incorrectly, the student is presented with an item of lower difficulty. The participants for his study were students located in rural eastern Pennsylvania (grades 1-4) who participated in fall, winter, and spring assessment periods within their district (Shapiro, 2012). However, this literature review singles out data related to the PSSA, CAT (i.e., STAR-Math), and M-CBM for students in grades 3 and 4 in order to show findings related to high stakes testing (n = 83 and 92, respectively) (Shapiro, 2012). STAR-Math had the strongest statistically significant correlations with the PSSA, ranging from .58 to .63, while M-CBM correlations ranged in the low range with some scores being not significant from .12 to .41 (Shapiro, 2012, p. 302). Separately, the researcher addressed the relationship between the M-CBM and STAR-Math. According to Shapiro (2012) "[t]hese data suggest that although the CBM

and CAT math measures were related to each other, they may be assessing somewhat different domains of mathematical knowledge" (p. 304). Therefore, the M-CBM does not have high statistical predictive validity of the PSSA.

Demographic Research as it Relates to Achievement/Standardized Assessments

Student populations are different in every school system. They vary in a multitude of demographic variables. However, every student, when they reach certain point in their educational career, is required to take assessments mandated by No Child Left Behind.

In relation to the Missouri Assessment Program (MAP), Hacker (2001) reported in the *St. Louis Post Dispatch* concern regarding the achievement gaps between African American students and other racial groups in Missouri and Illinois. In Missouri, she reported that third grade communication arts scores (between the years 1998 and 2001) for African American students rose 5.4 percentage points, Caucasian students rose 1.9 percentage points, and Asian students fell .3 percentage points (Hacker, 2001). Since African American student percentages rose more than Caucasian scores, the gap between these subgroups closed as scores got closer together on the assessments. In Illinois, the gap was also present. African American students performed poorly on the Illinois Standards Achievement Test in the year 2001, having the lowest demographic score on every grade level and subject tested (Hacker, 2001). The report did not show the relationship of the 2001 scores in Illinois and other assessment years. However, in a separate article, it was reported that this assessment was biased against minorities as all characters utilized within the assessment were portrayed as Caucasian (Parker, 2001).

Character portrayal within assessment items is one area where bias may occur. This happens when assessment items are written with all Caucasian characters.

The current study did not separate students from different racial subgroups. Therefore, included in this group are English Language Learners (ELL): students who come from bilingual home environments. The assessment data for these students may vary from students who have English as their native language. Abedi and Dietel (2004) summarized data for tenth grade students within the Boston Public Schools, as the 1998 Massachusetts Comprehensive Assessment System (MCAS) was included. ELL students performed significantly lower than their non-ELL peers, with only 7% of ELL students reaching the proficient or above levels on the English Language Arts portion of the assessment (Abedi & Dietel, 2004). Therefore, students who are a part of the ELL subgroup have been shown to attain lower scores on standardized assessments than their peers.

Harris (2007) studied the nation's schools in relation to performance categories, SES, and minority variables. His two research models showed first how high/low poverty schools performed on achievement tests and second how high poverty, high minority/low poverty, and low minority schools performed on achievement tests (Harris, 2007). The separate models are important since not all high poverty schools are in high minority areas and not all low poverty schools are in low minority areas. Therefore, by separating the data, different conclusions came forth. Harris (2007) found that of 40,830 low poverty schools (less than 50% of students eligible for free and reduced lunch), 54% were high performing; whereas, of 21,234 high poverty schools, 16% were high performing. The results showed that of these two groups, low poverty schools were more

likely to outperform schools that have a high poverty rate. In fact, "low-poverty schools are 22 times more likely to be high performing than high-poverty schools" (Harris, 2007, p. 389). If this is reduced to the lowest denominator, then it can be stated that a low poverty student is 22 times more likely to be high performing than a high poverty student. Additionally, low poverty, low minority schools (minority students making up less than 50% of the student population, n = 38,104) had a high performance rate of 54%; whereas, high poverty, high minority schools (n = 12,869) had a high performance rate of 10% (Harris, 2007). Harris (2007) found "Low minority schools are 89% more likely to reach this [i.e. high] performance level compared with their high-poverty, high-minority counterparts" (p. 389). In other words, a student who is low poverty, low minority is 89% more likely to reach a high performance level than a student who is high poverty, high minority.

Reardon and Galindo (2009) were concerned with a different racial achievement gap, that which affects Hispanic students in the areas of math and reading in elementary grades. Their student sample was from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K), sponsored by the National Center for Educational Statistics. Reardon and Galindo (2009) reported "The ECLS-K contains data [reported by school administration and family sources] on a nationally representative sample of approximately 21,400 students from the kindergarten class of 1998-1999" (p. 858). This sample was, therefore, a mirror of the national student population for the school year of 1998-1999. The researchers studied the Hispanic and Caucasian students included in this population to monitor the achievement gap as they progressed from kindergarten to fifth grade (Reardon & Galindo, 2009). It was important to demonstrate if a gap was present

and shows how the two separate racial groups performed on standardized assessments throughout the nation. The researchers found that Hispanic students in kindergarten were significantly lower in math and reading skills than were their Caucasian peers (math skills gap was between .77 and .92 standard deviations below, reading skills gap was between .51 and .61 standard deviations below) (Reardon & Galindo, 2009, p. 865). Data analysis at the end of fifth grade, for these same students, represented a math gap of between .60 and .50 standard deviations below and a gap in reading was between .45 and .38 standard deviations below their Caucasian peers (Reardon & Galindo, 2009, p. 866). Therefore, as these students traveled through the grade levels, they were able to narrow, but not eliminate, the gap between racial groups in reading and math. Although the study did not focus on African American students, the researchers stated that the gap of African American students compared to other racial backgrounds widens from kindergarten to fifth grade (Reardon & Galindo, 2009).

Hardegree (2012) researched if gender would account for student mean score differences on the Georgia Criterion Referenced Competency Test (GCRCT) for fifth grade students. Through an analysis of variance, the researcher found statistically significant differences between gender groups with p values that were less than 0.0001 within the models (Hardegree, 2012). Therefore, the researcher found that gender differences had an effect on the mean scores that students received on the GCRCT.

Summary

Formative and summative assessments are important means for determining student growth within school systems. Student populations react differently to the two

forms of assessments. This chapter provided a review of the literature for CBM (formative assessments) and high stakes testing (summative assessments). Chapter three includes the methodology employed in this study.

Chapter Three

Methods

The purpose of this study was to determine which independent variables (CBM scores, gender, SES, and race) were the best predictors of scores on the MAP. Chapter three describes the methodology used while conducting this study. This chapter includes a description of the research design, population and sample, instrumentation, procedures followed during data collection, data analysis and hypothesis testing, and the limitations of the study.

Research Design

This study employed a quantitative research design using archival data. Due to the nature of this study, multiple regression models were developed to address the research questions. "Multiple regression is a prediction equation that determines the correlation between a combination of two or more predictor variables and a criterion variable" (Lunenburg & Irby, 2008, p. 80). The goal of a stepwise multiple regression model is to identify the independent variable (i.e., predictor variable) combination that is the most parsimonious in predicting the dependent variable (i.e., criterion variable). Within this study, the independent variables were the fall, winter, and spring CBM assessment scores for math (M-CBM grades 3-5) and communication arts (R-CBM grade 3 and Maze CBM grades 4 and 5), and the demographic data of the students (gender, SES, and race). The dependent variables were the MAP assessment scores for communication arts and math for students in grades 3-5.

Population and Sample

During the 2010-2011 school year, 4,308 third through fifth grade students attended the NKCS (NKCS, 2010b). Of these students, 1,415 students attended third grade, 1,461 students attended fourth grade, and 1,432 students attended fifth grade (NKCS, 2010b). Table 1 (see chapter one) provided detailed demographic information about the population.

The sample for this study was a subset from the total population of third through fifth grade students who attended NKCS during the 2010-2011 school year. Students in the population who did not participate in all assessments did not qualify for the study. Each participant was administered reading and math CBM assessments in the fall, winter, and spring, and the MAP in the spring.

Sampling Procedures

Participants were selected from the third, fourth, and fifth grades of the NKCS. All students were eligible to participate in the study if they were enrolled in one of the 21 elementary schools during the assessment period of the 2010-2011 school year and all demographic variable data was present. If a student did not participate in all of the required assessments (fall CBMs, winter CBMs, spring CBMs, and the MAP) then they were excluded from the sample. In order to determine if students were eligible, student data was placed into the IBM® SPSS® Statistics Faculty Pack 21 for Windows. Students were excluded if their demographic data or any needed assessment data was missing.

CBM Instrumentation

The Curriculum Based Measurement (CBM) assessments are created and distributed by Pearson Publishing, under the PsychCorp brand that utilizes the

benchmarking software system (Pearson PsychCorp, 2010a). The purpose of Pearson PsychCorp (2010a) and the products that they provide are to assist teachers with progress monitoring of their students. This is done through three types of monitoring:

1. Benchmark – Assess all students three times per year for universal screening (early identification), general education progress monitoring, and AYP accountability.
2. Progress Monitor – Write individualized annual goals and monitor more frequently for those who need intensive instructional services.
3. Strategic Monitor – Monitor at-risk students monthly and evaluate the effectiveness of instructional changes. (para. 2)

The monitoring process scores that pertained to this study were the benchmark scores for Reading CBM (R-CBM), Maze CBM, and Math CBM (M-CBM).

R-CBM. The R-CBM assesses reading achievement components of oral reading fluency through a one-minute timed reading passage (Pearson PsychCorp, 2010c; Steckler, 2006). According to Pearson PsychCorp (2010c), "there are over 30 equivalent passages for each grade level" (para. 3). Hintze and Silberglitt (2005) indicated that the "passages were purposely developed with controlled vocabulary and difficulty and written by authors familiar with the teaching of reading and how students learn to read across a variety of types of literature" (p. 7). Of the published passages, the school district utilizes consistent passages for use in the fall, winter, and spring benchmarks across the district in order to standardize the assessments for all students. Students are given three passages at each benchmark and are given a median score. The staff members receive training and materials on the proper administration and scoring

procedures for the assessment; this standardization allows for comparison of student data immediately, and through time (Hintze & Silberglitt, 2005). Within the R-CBM administration process, a staff member works one-on-one with a student. Each has their own copy of the reading passage. The examiner copy is different from the student copy; it is marked with numbers for quick scoring (Shinn & Shinn, 2002b, p. 11). According to the AIMSweb Training Workbook, the passages are generally 250-300 words and begin with an introductory sentence (Shinn & Shinn, 2002b, p. 10). The test administrator, due to the standardized nature of the assessment, reads students uniform directions on how to take the assessment.

Measurement. The scorer is accountable for the words read correctly (WRC) and the number of errors. Therefore, as the student reads the passage orally, the assessor will mark on the examiner copy the WRC and the errors. The score that a student receives is then placed into the Pearson PsychCorp AIMSweb data system for NKCS with two parts: the WRC and the errors that the student read within the time allotted.

The technical information for the R-CBM reading passages differs from grade to grade. The information for grade 3 is provided as an example. The standardized passages, housed in AIMsweb, are used for grade 3 year after year within the North Kansas City Schools. Howe and Shinn (2002) detailed the technical information for these passages. Figure 1 shows the grade level standard reading passage information for grade 3 taken from the *Standard Reading Assessment Passages (RAPs) for General Outcome Measurement* (Howe & Shinn, 2002). The 33 passages read include the three standard benchmark assessments (including the fall, winter, and spring that are utilized

for the current study) and the 30 assessments which are used for progress monitoring throughout the year.

Number of Passages Read	33.00
Mean Words Read Correct (WRC)[a]	107.60
Standard Deviation[b]	28.10
Standard Error Measurement	10.50
Reliability[c]	.86

Figure 1. Technical Information for the Grade 3 Standard Reading Passages. Adapted from "Standard Reading Assessment Passages (RAPs) for General Outcome Measurement," by K. B. Howe & M. M. Shinn, 2002, p. 8.

[a]Average of RAP grade mean. [b]Average of RAP grade standard deviations. [c]Average of RAP grade correlations (alternate-form reliability).

For the purposes of this study, the yearly fall, winter, and spring standard benchmark assessments were utilized. Table 4 shows these specific assessments' technical information for grade 3. However, this table also includes the technical information for the third benchmarking session and the mean of all three passages' information as found by Howe and Shinn (2002).

Table 4

Technical Information for Grade 3 Standard Benchmark Reading Assessment Passages

Passage	Mean WRC	Standard Deviation	Alternate Form Reliability[a]	Lexile
1	107.1	29.4	.85	630
2	107.2	31.7	.83	460
3	109.1	28.2	.87	570
Mean	107.8	29.8	.85	

Note. Adapted from "Standard Reading Assessment Passages (RAPs) for General Outcome Measurement," by K. B. Howe & M. M. Shinn, 2002, p. 8.

[a]Mean correlation for each alternate form RAP.

Reliability and validity. Reliability is important for data collection procedures because it ensures that studies can be repeated with similar results. The assessment passages used for grade 3 provide adequate evidence that alternate form reliability are at high rates with a mean of .85 (Howe & Shinn, 2002). In addition to the reliability of an assessment measure, validity is also important for showing the accuracy of this assessment. "Validity is the degree to which an instrument measures what it purports to measure" (Lunenburg & Irby, 2008, p. 181). Research on R-CBM has demonstrated that timed measures, which count the number of correctly read words in 1 minute "is an excellent measure of general reading proficiency, including reading comprehension" (Hamilton & Shinn, 2003, p. 228). Researchers from the National Center for Response to Intervention (2010) found convincing evidence for criterion-related validity for third grade students at a coefficient of .72 for a sample of 78. Therefore, the R-CBM is a reliable and valid assessment.

Maze CBM. Along with the R-CBM, Pearson and PsychCorp developed the Maze CBM as an alternative measurement of reading ability to the R-CBM (Pearson PsychCorp, 2010b). The Maze is a multiple choice cloze task where the first sentence of a 150-400 word reading passage is presented intact; then, every seventh word is removed and the students must choose the correct multiple choice answer (Pearson PsychCorp, 2010b; Shinn & Shinn, 2002a, p. 8). Students have a choice of three words to complete the sentences: one correct word to complete the sentence, one distractor of the same type (noun, verb, etc.), and one word that is randomly selected from the text passage that is not of the same type (Shinn & Shinn, 2002a).

When trained teachers administer this assessment, the whole class is able to take the assessment at one time. However, due to the standardized nature of this assessment, the directions must be the same within all settings. The Maze is a timed assessment; students are given three minutes to read silently the passage given to them and circle the word that they think is correct (Shinn & Shinn, 2002a).

Measurement. In order to score the assessment, the assessor looks for the number of items marked correctly. The correct items and errors (which will vary due to the reading skills of each student) are counted and recorded (Shinn & Shinn, 2002a). The number of correct responses over the number of errors gives the administrator the student's score.

***Reliability and validity*.** Research has demonstrated support of reliability in the forms of alternate form and split-half scores for grades 1-7. Table 5 shows the statistical data to support the reliability of the performance level score in each instance.

Table 5

Reliability of the Maze CBM

Type of Reliability	*n*	Range	*Mdn*
Alternate Form	102	.66 to .91	.89
Split Half	102	.91 to .97	.95

Note. Adapted from "Progress Monitoring Tools," by National Center for Response to Intervention, 2010, para. 14.

There is also evidence of the validity of the Maze. Criterion validity for this assessment as found by the National Center for Response to Intervention (2010) for grades 1-7 was determined through the Comprehensive Reading Assessment Battery: WRC (n = 111). Scores ranged from .71 to .95 with a median of .83. National Center for Response to Intervention (2010) results support that the Maze is a reliable and valid assessment of reading.

M-CBM. Pearson Publishing created and distributed the M-CBM under the PsychCorp brand that utilizes the benchmarking software system (Pearson PsychCorp, 2010a). The purpose of the assessment is to discover a student's mathematical computation skills. This particular assessment is utilized for students in grades 1-6 or for grade levels higher than grade 6 when students fall within the lower grade level skill sets (Shinn, 2004).

The assessment is administered individually or class-wide; students in grades 1-3 are given two minutes to complete the assessment and students in grades 4-6 have four minutes to complete the assessment (Pearson PsychCorp, 2010d). The assessment that each individual student completes consists of a 2-page front and back probe (individual assessment page) to complete during the timeframe allotted (Pearson PsychCorp, 2010d).

The students write their answers on the probe page that they are given. According to Pearson PsychCorp (2010d), each probe contains the items that are created for the students within the developmental level of each grade level. In order to maintain standardized administrative procedures for the assessment, teachers and staff are trained in the administration of the assessment probes.

Measurement. In order to score the M-CBM, teachers are provided with the information on the correct method for scoring the assessment. The scoring procedures require tabulation for each digit correct within an answer; in addition, partial credit is a possibility for each problem that a student attempts to complete (Pearson PsychCorp, 2010d).

Reliability and validity. Evidence has been found for the reliability and validity of the M-CBM. "Scientific-based research has shown that having students write answers to grade-level computational story problems for 2-4 minutes is a reliable and valid general outcome measure of general mathematics computation for typically achieving students through Grade 6" (Shinn, 2004, p. 3). In fact, Thurber, Shinn, Smolkowski (2002) found within a study of 207 fourth graders that interscorer agreement reliability was at .83 and found alternate form reliability to be .91.

MAP Instrumentation

The Missouri Assessment Program (MAP) monitors student progress towards the mastery of the Missouri State Show Me Standards (see chapter one). The 2010-2011 assessment was composed of two item formats: constructed response and multiple-choice (see definition of terms in chapter one). The MAP includes items created by the Missouri State Assessment authors and test items chosen from the *TerraNova* Survey (DESE,

2011c), which is a product of CTB/McGraw-Hill, a publishing company. *TerraNova* Survey is an abbreviated version of the Complete Battery and provides a general measure of achievement in a minimum amount of testing time. The survey generates norm-referenced achievement scores to measure students' academic levels in different curriculum areas (CTB/McGraw-Hill, 2010b).

According to DESE (2011b), MAP subtests assess three areas of communication arts standards and five areas of mathematics standards. Communication arts subtests include writing standard English, reading, and formal writing. The area of writing standard English focuses on grammar, spelling, and sentence/paragraph structure. The area of reading has items focused on fiction and non-fiction texts. Formal writing items focus on audience awareness, writing prompts, and writing formats in letters and graphic organizers.

Within the mathematics MAP assessment, the standard subtests are number and operations, geometric and spatial relationships (including measurement), data and probability, and algebraic relationships (DESE, 2011b). Number, operation, and algebraic relationship test items include questions based on number sense, addition, subtraction, multiplication, and division applications. Geometric and spatial relationship (including measurement) problems focus on length, area, volume, and applications concerning shapes. When questioned on data and probability, students are shown graphs and figures and asked to evaluate them for various statistical facts.

Measurement. For each area assessed on the MAP, students receive an achievement level of *Below Basic, Basic, Proficient,* and *Advanced* in relation to their score, as shown in Table 6. The achievement level that relates to each score range is a

label, which the publishers and Missouri school systems use in order to group students by achievement level. This study utilized the score within the multiple regression models to find if a student's CBM assessment score predicted the achievement level that the student acquired on the MAP assessment (respective to curricular area).

Table 6

MAP Scores in Relation to Achievement Level

Grade Level	Curriculum Area	Achievement Level	Score Range
Grade 3	Communication Arts	Below Basic	455 – 591
		Basic	592 – 647
		Proficient	648 – 672
		Advanced	673 – 790
	Math	Below Basic	450 – 567
		Basic	568 – 627
		Proficient	628 – 666
		Advanced	667 - 780
Grade 4	Communication Arts	Below Basic	470 - 611
		Basic	612 - 661
		Proficient	662 - 690
		Advanced	691 - 820
	Math	Below Basic	465 – 595
		Basic	596 – 650
		Proficient	651 – 687
		Advanced	668 – 805
Grade 5	Communication Arts	Below Basic	485 – 624
		Basic	625 - 674
		Proficient	675 – 701
		Advanced	702 - 840
	Math	Below Basic	480 – 604
		Basic	605 – 667
		Proficient	668 – 705
		Advanced	706 - 830

Note. Adapted from "Missouri Achievement Program: Guide to Interpreting Your Results," by CTB/McGraw-Hill, 2011, p. 8.

The scoring procedures vary by problem type. The multiple-choice assessments are scored by machine (DESE, 2011c). Stiggins et al. (2006) stated "a key to successful

use of extended written response assessment is the clear articulation of appropriate evaluation criteria by which to judge the quality of student responses" (p. 177). Thus, the scoring of the constructed response test items has clearly defined procedures. Scoring begins with a rigorous screening and interviewing process for hand scorers of the constructed response section (CTB McGraw-Hill, 2010a). Once selected, the hand-scorers meet criteria by attending a training meeting where rubrics and previously field-tested Missouri operational test items are introduced and scored (CTB McGraw-Hill, 2010a).

***Reliability and validity*.** The scoring procedures for the MAP assessments are of vital importance to ensure confidentiality, reliability, and validity of scores. The state of Missouri has put into place practices of scoring which adhere to the 1999 standards set by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) in the *Standards for Educational and Psychological Testing* (CTB/McGraw-Hill, 2010a).

During the process of hand scoring student assessments, several actions were in place to secure reliability and validity of the scoring process:

- Daily accuracy checks were performed to monitor the validity through the grading of "validity sets." These are pre-scored assessments that are graded by the administration and then re-graded by the scores. The administration checks for matching grades to demonstrate validity.
- A read-behind process to insure inter-rater reliability. Five percent of assessments were scored by two handscorers to insure inter-rater reliability. All tested areas "show good inter-rater agreement… raters demonstrated at

least 84% perfect and adjacent agreement for Communication Arts items… and at or above 98% perfect or adjacent agreement for all Mathematic items."

- Recalibration of raters – if a rater continued to drift then he/she was released. (CTB McGraw-Hill, 2010a, pp. 62-63)

Other measures were taken to ensure reliability and validity of the MAP assessment. CTB McGraw-Hill (2010a) developed the assessment following the AERA, APA, and NCME standards, which stress creating and maintaining tests of the highest quality. They "calculated the reliability of each MAP test in a variety of ways: reliability of raw scores, overall standard error of measurement [SEM], IRT [item response theory]-based conditional standard error of measurement and decision consistency of achievement level classifications" (CTB McGraw-Hill, 2010a, p. 132).

The raw score reliability "was evaluated using Cronbach's coefficient alpha, which is a lower-bound estimate of test reliability" (CTB McGraw-Hill, 2010a, p. 132). In order to judge the reliability of a test, the coefficient is calculated; these scores range from 0-1. Depending on the data being utilized, different coefficients are considered an acceptable score: research purposes require a minimum score of .70; for tests of moderate length, a minimum score of .80 is required; and for clinical testing purposes, a score of .90 is needed (CTB McGraw-Hill, 2010a, p. 133; Johnson & Christensen, 2008, p. 149). Table 7 shows the reliability data for the communication arts subtest. The coefficients provide strong evidence for the reliability of the test with coefficients greater than .90 (Johnson & Christensen, 2008, p. 149).

Table 7

Reliability Coefficient Data for Communications Arts

Grade	Number of Items	Cronbach's α
3	56	0.91
4	58	0.93
5	56	0.91

Note. Adapted from "Missouri Grade Level Assessment Technical Report 2010," by CTB McGraw-Hill, 2010a, p. 142.

Table 8 shows the reliability data for the mathematics subtest. The coefficients provide strong evidence for the reliability of the test with coefficients above .90 (Johnson & Christensen, 2008, p. 149).

Table 8

Reliability Coefficient Data for Mathematics

Grade	Number of Items	Cronbach's α
3	55	0.92
4	62	0.92
5	58	0.91

Note. Adapted from "Missouri Grade Level Assessment Technical Report 2010," by CTB McGraw-Hill, 2010a, p. 142.

Finally, it is important to note that the classifications of performance levels are reliable for student placement in *Below Basic, Basic, Proficient,* and *Advanced*. Classification accuracy and classification consistency are used to determine the reliability of these levels. Classification accuracy refers to the rate at which the student's true score (or numerical value) is related to the student's

placement. Classification accuracy on the MAP assessment is at or above .90 (CTB McGraw-Hill, 2010a, p. 137). Classification consistency is defined as how consistent two parallel assessment forms accurately determine the same student placement levels. For the MAP, classification consistency is at or above .87; therefore, "[t]hese results suggest that consistent and accurate performance level classifications are being made for students in Missouri based on the MAP" (CTB McGraw-Hill, 2010a, p. 137).

Chapter 6 within the *Missouri Grade Level Technical Report* presents and summarizes the validity of the assessment or that "the test items, as well as the overall test, are functioning appropriately" (CTB McGraw-Hill, 2010a, p. 78). Evidence of validity is approximately .75, indicative of MAP test scores accurately measuring the appropriate content (CTB McGraw-Hill, 2010a, p. 71).

Data Collection Procedures

Each year the Missouri DESE has an open enrollment period for districts to enroll their students for the MAP. This online enrollment period was from November 29–December 10, 2010 for the 2010-2011 school year (VanDeZande, 2010). The assessment Test Coordinator Packages arrived in Missouri school districts on February 23, 2011 (VanDeZande, 2010). The materials for the assessment arrived in the school districts on March 11, 2011 (VanDeZande, 2010). School districts administered the tests during the state's testing window from March 28–April 22, 2011; districts then contacted CTB/McGraw Hill to pick up the assessments by April 25, 2011 (VanDeZande, 2010). After the assessment collection by CTB/McGraw Hill, the assessment scoring took place at the state level and participating districts then received the released data when the analysis was completed.

During the 2010-2011 school year, NKCS utilized three assessment windows to maintain standardization of the CBM assessments, as illustrated in Table 9. CBM administration occurred as subtests. The subtests given to NKCS students were the Oral Reading Fluency (R-CBM), Maze, and Math CBM (M-CBM). Grades 3-5 were administered the MAP assessment for communication arts and math; in addition, they each took the Math CBM. However, the R-CBM was administered at grade 3 and the Maze was administered to grades 4 and 5.

Table 9

2010-2011 Elementary Testing Windows for NKCS

Assessment	Assessment Window
Curriculum-Based Measurements (R-CBM, Maze, and Math)	September 1-15 January 10-21 May 2-13
Missouri Assessment Program (MAP) (Communication Arts and Math)	April 4-15

Note. Adapted from "Elementary Assessment Windows," by NKCS, 2010a, pp. 1, 3.

Teachers scheduled the assessments within their class day and scored them during their planning time. District personnel then entered the CBM scores into the Pearson PsychCorp AIMSweb computer system (L. Friesen, personal communication, November 27, 2011).

Additionally, protocols for privacy, confidentiality, and student rights are a concern when conducting research. Therefore, prior to data collection, strict guidelines were in place to maintain these rights. First, Baker University's Institutional Review Board (IRB) reviewed and accepted a proposal of research concerning this study (see

Appendix B). Next, the Research Committee at NKCS reviewed the IRB findings along with their district data collection request form (see Appendix C). When both Baker University (see Appendix D) and NKCS personnel (see Appendix E) had approved the study, the researcher was able to move forward with data collection.

Numerical and coded data of the R-CBM, Maze CBM, M-CBM, and the MAP assessments were collected. The school district removed the participants' names from the data set of the study to protect their identities. The numerical data was input into IBM® SPSS® Statistics Faculty Pack 21 for Windows for analysis.

Data Analysis and Hypothesis Testing

The study examined the following research questions to determine which combination of variables best predicts student scores on the MAP assessments. The research questions provide the basis for the data analysis.

RQ1: What combination of variables (fall Oral Reading Fluency CBM score, winter Oral Reading Fluency CBM score, spring Oral Reading Fluency CBM, gender, SES, and race) best predicts student scores on the third grade MAP assessment in communication arts?

H1: The best combination of variables for predicting student scores on the third grade MAP communication arts assessment includes fall Oral Reading Fluency CBM score, winter Oral Reading Fluency CBM score, spring Oral Reading Fluency CBM, gender, SES, and race.

RQ2: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in communication arts?

H2: The best combination of variables for predicting student scores on the fourth grade MAP communication arts assessment includes fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race.

RQ3: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in communication arts?

H3: The best combination of variables for predicting student scores on the fifth grade MAP communication arts assessment includes fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race.

RQ4: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the third grade MAP assessment in math?

H4: The best combination of variables for predicting student scores on the third grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

RQ5: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in math?

H5: The best combination of variables for predicting student scores on the fourth grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

RQ6: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in math?

H6: The best combination of variables for predicting student scores on the fifth grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

Multiple regression analyses were used to find the combination of variables that best predicted MAP scores for each grade level. A stepwise regression model was conducted for each of the grade levels (3-5) and subtests (communication arts and mathematics). Correlations were examined for the strength of the associations between each of the predictor variables and the MAP assessment scores, and the variables were tested for their contributions to the final models.

Categorical variables were dummy coded with two categories and were added to the multiple regression models. Dummy codes are artificial numerical values assigned to the categories, coded as 0 and 1. These numbers provide a "classification, not a magnitude, for the categorical predictor" (Agresti & Finlay, 2009, p. 379). Therefore, the demographic information was coded in the order that the categorical variables were included in the multiple regression models. For each grade level, gender was coded male = 0 and female = 1, SES was coded free or reduced status = 0 and full pay = 1, and race was coded non-white = 0 and white = 1. The full listings of predictor variables for third grade were fall, winter, and spring Oral Reading Fluency scores, fall, winter, and spring M-CBM scores, gender, SES, and race. For fourth grade students, the predictor variables were fall, winter, and spring Maze CBM scores, fall, winter, and spring M-CBM scores,

gender, SES, and race. Predictor variables for fifth grade students were fall, winter, and spring Maze CBM scores, fall, winter, and spring M-CBM scores, gender, SES, and race.

Limitations

In all assessment administration, there is an element of subjectivity to the assessment. Every effort was taken to train staff members to administer the assessments in a standardized nature. However, since the same person does not administer the assessments, some variance may occur in statement of directions, the time administration, and the collection procedures of the assessments. These variances are limitations to the results of this study.

An additional limitation for the study involved the sample. This study utilized only one year of data. The MAP assessment does vary from year to year, whereas, the CBMs do not.

Summary

This chapter gave a detailed account of the research design, the population and sample, the MAP and CBM instrumentation, the data analysis and hypothesis testing, and limitations of the study. Chapter four includes the results of the study.

Chapter Four

Results

The purpose of this study was to determine if student performance on Curriculum Based Measurement (CBM) predicted 3rd-5th grade student success on the Missouri Assessment Program (MAP) assessments in the areas of communication arts and math. An additional purpose was to determine if demographic variables were also predictors of the MAP scores. This chapter includes the descriptive statistics and results of the hypothesis testing.

Descriptive Statistics

The potential participants in this study were enrolled in the third, fourth, and fifth grades in the NKCS during the 2010-2011 school year (n = 4,232). However, not all of these students were included in the study. If a student did not have reported information for each of the variables (fall, winter, and spring R-CBM scores or Maze scores, MAP communication arts scores, gender, SES, and race) necessary for research questions 1-3, they were excluded from this portion of the study. Additionally, if a student did not have reported information for each of the variables (fall, winter, and spring M-CBM scores, MAP mathematics scores, gender, SES, and race) necessary for research questions 4-6, they were excluded from this portion of the study. If a student had the necessary reading scores, but not the necessary math scores, they were included in the sample for the reading portion of the study, but not the math portion; the reverse was also true. No student was considered as a part of the sample for this study if the necessary demographic variables were not included. The final samples included 1,274 students in the third grade,

1,261 in the fourth grade, and 1,270 in the fifth grade. Table 10 illustrates the number of students included in each of the samples to address the intended research question.

Table 10

Sample Sizes Per Research Question

Research Question	Assessment	Grade	*n*
1	R-CBM	3	886
2	Maze	4	1055
3	Maze	5	1020
4	M-CBM	3	450
5	M-CBM	4	520
6	M-CBM	5	377

The sample sizes for each research question differ due to only including those students who had assessment scores for each of the three test administrations for reading or math. Students overlapped samples if the student's data included scores from all six test administrations.

Three demographic variables were required for a student to be included in a sample: gender, SES, and race. In regards to SES, 1,954 students were registered as full pay lunch status, while 1,851 students were registered with a free or reduced lunch status. The total sample included 1,842 female students and 1,963 male students, and 1,309 non-white students and 2,496 white students. Table 11 shows the demographic data disaggregated by research question.

Table 11

Demographic Data by Research Question

Demographic	RQ 1	RQ2	RQ3	RQ4	RQ5	RQ6
SES						
Full Pay	484	577	560	259	273	207
Free/Reduced	402	478	460	191	247	170
Gender						
Male	300	348	347	147	175	123
Female	586	707	673	303	345	254
Race						
White	426	533	493	229	254	191
Non-White	460	522	527	221	266	186

Demographic data was included in the regression models to determine the extent that the variables were predictors of MAP scores.

Hypothesis Testing

This section includes the results of the hypothesis testing. Multiple regression models using stepwise methods were used to address each of the six research questions for this study. Included in each hypothesis discussion is the combination of variables that produced the most parsimonious regression model. The level of significance for each multiple regression model was set at .05.

RQ1: What combination of variables (fall Oral Reading Fluency CBM score, winter Oral Reading Fluency CBM score, spring Oral Reading Fluency CBM, gender,

SES, and race) best predicts student scores on the third grade MAP assessment in communication arts?

H1: The best combination of variables for predicting student scores on the third grade MAP assessment in communication arts includes fall Oral Reading Fluency CBM score, winter Oral Reading Fluency CBM score, spring Oral Reading Fluency CBM, gender, SES, and race.

The independent variables used in this analysis included the fall, winter, and spring Oral Reading Fluency CBM (R-CBM) scores, gender, SES, and race. For research question one, all variables were reviewed for the third grade students of the North Kansas City Schools. The dependent variable for research question one was the MAP communication arts assessment for third grade students. Students, with reported demographic data, who participated in the communication arts MAP assessment and each of the R-CBMs, were included in this analysis (n = 886).

A multiple regression analysis using a stepwise method was conducted to test hypothesis one. The most parsimonious model identified for predicting MAP communication arts scores for third grade included all variables, F = 159.095, df = 6, 879, $p < .001$. The variables accounted for 52.1% of the variability in communication arts scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (B) was tested using a t test (see Table 12). The regression equation for predicting MAP communication arts scores for third grade students was: 3rd Grade Communication Arts MAP Score = .162(R-CBM Fall Score) +.25(R-CBM Winter Score) + .181(R-CBM Spring Score) + 4.795(Gender) + 4.916(SES) + 5.257(Race) + 569.08.

Table 12

Slope Coefficients and t *Tests for Research Question One*

Variable	*B*	*t*	*p*
CBM			
Fall	.162	2.873	.004
Winter	.250	3.552	< .001
Spring	.181	2.908	.004
SES	4.916	2.870	.004
Gender	4.795	2.974	.003
Race	5.257	2.981	.003

Winter R-CBM scores had the strongest relationship to MAP communication arts scores (r = .702); fall and spring scores had moderately strong relationships (r = .684 and .691, respectively). Of the three demographic variables, SES has the strongest (although weak) relationship to the MAP communication arts scores (r = .226), followed by race (r = .151) and gender (r = .124). All correlations were significant at $p < .001$. All of the variables were predictors of MAP communication arts scores, which supports hypothesis one.

RQ2: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in communication arts?

H2: The best combination of variables for predicting student scores on the fourth grade MAP communication arts assessment includes fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race.

The independent variables used in this analysis included the fall, winter, and spring Maze CBM scores, gender, SES, and race. The Maze assessment scores (fall, winter, and spring) show silent reading fluency and comprehension of a written passage. For research question two, all variables were reviewed for the fourth grade students of the NKCS. The dependent variable for research question two was the MAP communication arts assessment for fourth grade students. Students, with reported demographic data, who participated in the communication arts MAP assessment and each of the Maze CBMs, were included in this analysis (n = 1055).

A multiple regression analysis using stepwise methods was conducted to test hypothesis two. The most parsimonious model identified for predicting MAP communication arts scores for fourth grade included fall, winter, and spring Maze scores and SES, F = 247.602, df = 4, 1050, p < .001. The variables accounted for 48.5% of the variability in communication arts scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (B) was tested using a t test (see Table 13). Gender and race were not contributors to the final model; therefore, they were excluded. The regression equation for predicting MAP communication arts scores for fourth grade students was: 4^{th} Grade Communication Arts MAP Score = 1.009(Maze CBM Fall Score) + 1.101(Maze CBM Winter Score) + 1.207(Maze CBM Spring Score) + 5.304(SES) + 597.303.

Table 13

Slope Coefficients and t *Tests for Research Question Two*

Variable	*B*	*t*	*p*
Maze			
Fall	1.009	5.271	< .001
Winter	1.101	6.906	< .001
Spring	1.207	8.175	< .001
SES	5.304	3.265	.001

Fall, winter, and spring Maze scores (r = .594, .641, and .641, respectively) had moderate relationships with MAP communication arts scores, and SES (r = .278) had a weak relationship with MAP communication arts scores. All correlations were significant at p < .001. Four of the variables were predictors of MAP communication arts scores, which marginally supports hypothesis two.

RQ3: What combination of variables (fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in communication arts?

H3: The best combination of independent variables for predicting student scores on the fifth grade MAP communication arts assessment includes fall Maze CBM score, winter Maze CBM score, spring Maze CBM score, gender, SES, and race.

The independent variables used in this analysis included the fall, winter, and spring Maze CBM scores, gender, SES, and race. The Maze assessment scores (fall, winter, and spring) show silent reading fluency and comprehension of a written passage. For research question three, all variables were reviewed for the fifth grade students of the

North Kansas City Schools. The dependent variable for research question three was the MAP communication arts assessment for fifth grade students. Students with reported demographic data, who participated in the communication arts MAP assessment and each of the Maze CBMs, were included in this analysis (n = 1020).

A multiple regression analysis using stepwise methods was conducted to test hypothesis three. The most parsimonious model identified for predicting MAP communication arts scores for fifth grade included fall, winter, and spring Maze scores and SES, F = 319.416, df = 4, 1015, p < .001. The variables accounted for 55.7% of the variability in communication arts scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (B) was tested using a t test (see Table 14). Gender and race were not contributors to the final model; therefore, they were excluded. The regression equation for predicting MAP communication arts scores for fifth grade students was: 5^{th} Grade Communication Arts MAP Score = 1.386(Maze CBM Fall Score) + .978(Maze CBM Winter Score) + .900(Maze CBM Spring Score) + 3.996(SES) + 601.277.

Table 14

Slope Coefficients and t *Tests for Research Question Three*

Variable	*B*	*t*	*p*
Maze			
Fall	1.386	8.665	< .001
Winter	.978	6.471	< .001
Spring	.900	6.202	< .001
SES	3.996	2.655	.008

Fall, winter, and spring Maze scores (r = .694, .690, and .674, respectively) had moderate relationships with MAP communication arts scores, and SES (r = .246) had a weak relationship with MAP communication arts scores. All correlations were significant at p < .001. Four of the variables were predictors of MAP communication arts scores, which marginally supports hypothesis three.

RQ4: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the third grade MAP assessment in math?

H4: The best combination of variables for predicting student scores on the third grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

The independent variables used in this analysis included the fall, winter, and spring Math CBM (M-CBM) scores, gender, SES, and race. The M-CBM assessment scores (fall, winter, and spring) show students' mathematics computation and fluency standardized scores. For research question four, all variables were reviewed for the third grade students of the NKCS. The dependent variable for research question four was the MAP math assessment for third grade students. Students with reported demographic variables, who participated in the math MAP assessment and each of the M-CBMs, were included in this analysis (n = 450).

A multiple regression analysis using stepwise methods was conducted to test hypothesis four. The most parsimonious model identified for predicting MAP math scores for third grade included fall, winter, and spring M-CBM scores and SES, F = 79.119, df = 4, 445, p < .001. The variables accounted for 41.6% of the variability in

math scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (*B*) was tested using a *t* test (see Table 15). Gender and race were not contributors to the final model; therefore, they were excluded. The regression equation for predicting MAP math scores for third grade students was: 3rd Grade Math MAP Score = .995(M-CBM Fall Score) + 2.027(M-CBM Winter Score) + 1.403(M-CBM Spring Score) + 7.22(SES) + 573.955.

Table 15

Slope Coefficients and t *Tests for Research Question Four*

Variable	*B*	*t*	*p*
M-CBM			
Fall	.995	2.369	.018
Winter	2.027	5.912	< .001
Spring	1.403	6.431	< .001
SES	7.220	2.550	.011

Fall, winter, and spring M-CBM scores (r = .480, .575, and .549, respectively) had moderate relationships with MAP math scores, and SES (r = .233) had a weak relationship with MAP math scores. All correlations were significant at $p < .001$. Four of the variables were predictors of MAP math scores, which marginally supports hypothesis four.

RQ5: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fourth grade MAP assessment in math?

H5: The best combination of variables for predicting student scores on the fourth grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

The independent variables used in this analysis included the fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race. The M-CBM assessment scores (fall, winter, and spring) show students' mathematics computation and fluency standardized scores. For research question five, all variables were reviewed for the fourth grade students of the NKCS. The dependent variable for research question five was the MAP math assessment for fourth grade students. Students with reported demographic data, who participated in the math MAP assessment and each of the M-CBMs, were included in this analysis (n = 520).

A multiple regression analysis using stepwise methods was conducted to test hypothesis five. The most parsimonious model identified for predicting MAP math scores for fourth grade included fall, winter, and spring M-CBM scores and SES, F = 89.924, df = 4, 515, $p < .001$. The variables accounted for 41.1% of the variability in math scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (B) was tested using a t test (see Table 16). Gender and race were not contributors to the final model; therefore, they were excluded. The regression equation for predicting MAP math scores for fourth grade students was: 4^{th} Grade Math MAP Score = 1.109(M-CBM Fall Score) + 1.11(M-CBM Winter Score) + .841(M-CBM Spring Score) + 6.037(SES) + 598.507.

Table 16

Slope Coefficients and t *Tests for Research Question Five*

Variable	*B*	*t*	*p*
M-CBM			
Fall	1.109	4.491	< .001
Winter	1.110	5.595	< .001
Spring	.841	5.437	< .001
SES	6.037	2.854	.004

Fall, winter, and spring M-CBM scores (r = .544, .551, and .528, respectively) had moderate relationships with MAP math scores, and SES (r = .280) had a weak relationship with MAP math scores. All correlations were significant at $p < .001$. Four of the variables were predictors of MAP math scores, which marginally supports hypothesis five.

RQ6: What combination of variables (fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race) best predicts student scores on the fifth grade MAP assessment in math?

H6: The best combination of independent variables for predicting student scores on the fifth grade MAP math assessment includes fall Math CBM score, winter Math CBM score, spring Math CBM score, gender, SES, and race.

The independent variables used in this analysis included the fall, winter, and spring Math CBM (M-CBM) scores, gender, SES, and race. The M-CBM assessment scores (fall, winter, and spring) show students' mathematics computation and fluency standardized scores. For research question six, all variables were reviewed for the fifth

grade students of the NKCS. The dependent variable for research question six was the MAP math assessment for fifth grade students. Students with reported demographic data, who participated in the math MAP assessment and each of the M-CBMs, were included in this analysis (n = 377).

A multiple regression analysis using stepwise methods was conducted to test hypothesis six. The most parsimonious model identified for predicting MAP math scores for fifth grade included fall, winter, and spring M-CBM scores, F = 132.764, df = 3, 373, $p < .001$. The variables accounted for 51.6% of the variability in math scores. Each of the variables in the model was tested for a significant contribution to the model. Each slope coefficient (B) was tested using a t test (see Table 17). SES, gender, and race were not contributors to the final model; therefore, they were excluded. The regression equation for predicting MAP math scores for fifth grade students was: 5^{th} Grade Math MAP Score = 1.957(M-CBM Fall Score) + 1.685(M-CBM Winter Score) + 1.589(M-CBM Spring Score) + 613.193.

Table 17

Slope Coefficients and t *Tests for Research Question Six*

Variable	*B*	*t*	*p*
M-CBM			
Fall	1.957	4.368	< .001
Winter	1.685	4.523	< .001
Spring	1.589	4.805	< .001

Fall, winter, and spring M-CBM scores (r = .638, .658, and .654, respectively) had moderate relationships with MAP math scores. All correlations were significant at $p <$

.001. Three of the variables were predictors of MAP math scores, which marginally supports hypothesis six.

Summary

This chapter included the descriptive statistics and results of hypothesis testing for this study. Multiple regression models were conducted to examine which combinations of variables best predicts communication arts and math MAP assessment scores. The data analyses show that the CBM scores have statistically significant relationships with their corresponding MAP assessments. However, the demographic variables did not have statistically significant relationships with the MAP assessments in all of the analyses. Chapter five includes an overview of the study, major findings, findings related to the literature, implications for action, recommendations for future research, and concluding remarks.

Chapter Five

Interpretation and Recommendations

Formative and summative assessments each have an important role within the educational system. Formative assessments help to guide instruction, while summative assessments show what a student has learned at the end of an instructional time frame. The purpose of this study was to determine if Curriculum Based Measurements (CBM), formative assessments, could predict student achievement on the MAP, a summative assessment. Chapter five includes a study summary, the findings of this study as they relate to the literature review, and the conclusions of the research.

Study Summary

The objective of this study was to determine if formative assessments (CBM) and demographic data (gender, SES, and race) were predictors of summative assessments (MAP). The results support the use of formative assessments as predictors of student achievement on the summative assessments. Within this section, an overview of the problem, purpose statement and research questions, review of the methodology, major findings, and findings related to literature are discussed.

Overview of the problem. Accountability within the school setting requires that teachers and administrators are able to show that students learn what they are required to learn. Within the state of Missouri, students participate in the Missouri Assessment Program (MAP) assessments in the areas of communication arts and math. In the elementary program, these summative assessments are administered once per year, in the spring, to students in third, fourth, and fifth grades. There is no way to predict how

students will perform on the MAP. Curriculum Based Measurements (CBM) are formative assessments used within the North Kansas City Schools to show student progress throughout the school year. The problem researched in this study was whether CBM scores and demographic data were predictors of students' MAP scores.

Purpose statement and research questions. The purpose of this study was to determine if formative CBM assessments were predictors of student performance on the MAP. An additional purpose was to determine if demographic variables were also predictors of student scores on the MAP. Six research questions were developed to determine if any of the independent variables (fall, winter, and spring reading and math CBM assessment scores, gender, SES, and race) were predictors of student performance on the respective dependent variables (communication arts and math MAP scores). The population and sample for this study included 3rd-5th grade students in the North Kansas City Schools (NKCS).

Review of the methodology. A quantitative research design was utilized in this study. Third through fifth grade students enrolled in NKCS during the 2010-2011 school year were the sample for this study. The Director of Research, Evaluation, and Accountability for the NKCS gathered the data utilized in this study after formal Institutional Review Boards for both NKCS and Baker University approved the study (see Appendices D and E).

The hypotheses in this study stated that a combination of all independent variables (fall, winter, and spring CBM scores, gender, SES, and race) were predictors of the dependent variables (MAP assessment scores). Data was collected and input into IBM® SPSS® Statistics Faculty Pack 21 for Windows for analyses. Multiple regression models

using stepwise methods were conducted; the correlations were analyzed for the strength of the relationship between each of the independent variables and the MAP assessment scores, and each variable was tested for its significant contribution to the model.

Major findings. Chapter four provided a detailed presentation of the results of each of the six research questions and hypotheses. A multiple regression model was used to address each of the research questions.

The evidence provided by the multiple regression model for research question one supported the first hypothesis. Fall, winter, and spring Oral Reading Fluency CBM scores, gender, SES, and race were strong predictors of third grade students' scores on the MAP communication arts assessment. The multiple regression results suggest that third grade students who have higher scores for the fall, winter, and spring R-CBM are more likely to have higher communication arts MAP scores. Further, female students are more likely to have higher MAP scores than male students; white students are more likely to have higher scores than non-white students; and students with full pay lunch status are more likely to have higher scores than students with free and reduced lunch status.

The evidence provided by the multiple regression model for research question two marginally supported the second hypothesis. Fall, winter, and spring Maze CBM scores and SES were strong predictors of fourth grade students' scores on the MAP communication arts assessment. Race and gender were not included in the final model as predictors of fourth grade MAP communication arts assessment scores. The multiple regression results suggest that fourth grade students, who have higher scores for the fall, winter, and spring Maze CBM, are more likely to have higher communication arts MAP

scores. Further, students with full pay lunch status are more likely to have higher MAP scores than students with free and reduced lunch status.

The evidence provided by the multiple regression model for research question three marginally supported the third hypothesis. Fall, winter, and spring Maze CBM scores and SES were strong predictors of fifth grade students' scores on the MAP communication arts assessment. Race and gender were not included in the final model as predictors of the fifth grade MAP communication arts assessment scores. The multiple regression results suggest that fifth grade students, who have higher scores for the fall, winter, and spring Maze CBM, are more likely to have higher communication arts MAP scores. Further, students with full pay lunch status are more likely to have higher MAP scores than students with free and reduced lunch status.

The evidence provided by the multiple regression model for research question four marginally supported the fourth hypothesis. Fall, winter, and spring M-CBM scores and SES were strong predictors of third grade students' scores on the MAP mathematics assessment scores. Race and gender were not included in the final model as predictors of third grade MAP mathematics assessment scores. The multiple regression results suggest that third grade students, who have higher scores for the fall, winter, and spring M-CBM, are more likely to have higher math MAP scores. Further, students with full pay lunch status are more likely to have higher MAP scores than students with free and reduced lunch status.

The evidence provided by the multiple regression model for research question five marginally supported the fifth hypothesis. Fall, winter, and spring M-CBM scores and SES were strong predictors of fourth grade students' scores on the MAP mathematics

assessment scores. Race and gender were not included in the final model as predictors of fourth grade MAP mathematics assessment scores. The multiple regression results suggest that fourth grade students, who have higher scores for the fall, winter, and spring M-CBM, are more likely to have higher math MAP scores. Further, students with full pay lunch status are more likely to have higher MAP scores than students with free and reduced lunch status.

The evidence provided by the multiple regression model for research question six marginally supported the sixth hypothesis. Fall, winter, and spring M-CBM scores were strong predictors of fifth grade students' scores on the MAP mathematics assessment scores. Race, gender, and SES were not included in the final model as predictors of fifth grade MAP mathematics MAP assessment scores. The multiple regression results suggest that fifth grade students, who have higher scores for the fall, winter, and spring M-CBM, are more likely to have higher math MAP scores.

Findings Related to the Literature

The goal of this study was to extend the current knowledge linking formative and summative assessments. Chapter two provided an extensive description of published research related to this study. This section relates the findings of this study to the research presented in chapter two.

Research question one asked what combination of variables (R-CBM scores, gender, SES, and race) would best predict student scores on the MAP assessment. Kranzler et al. (1999) found that there was no bias towards gender and race/ethnicity (of African American and Caucasian students) when correlating the R-CBM and the California Achievement Test in grade 3. The current study found that in grade 3, gender

and race (white and non-white) were predictors of student success on the MAP assessment, indicating that students who were female or white would have higher scores than students who were male or non-white. Although this is not a bias, it does give information related to how students will perform on the assessment when sorted by these subgroups.

Stage and Jacobsen (2001) researched the predictive relationship of the R-CBM and the Washington Assessment of Student Learning (WASL). Their findings showed that fall, winter, and spring fourth grade R-CBM scores could predict student achievement on the fourth grade WASL (Stage & Jacobsen, 2001). The current study also found the fall, winter, and spring R-CBM scores were strong predictors of the MAP assessment. Therefore, this study agrees with the findings of Stage and Jacobsen (2001) that R-CBM scores are predictors of state-developed summative assessments. However, the sample for the first research question in the current study included only third grade students.

McGlinchey and Hixson (2004) replicated the study of Stage and Jacobsen (2001) to find if the R-CBM could predict student achievement on the fourth grade Michigan Educational Assessment Program (MEAP) reading assessment. The findings of their study indicated that the R-CBM scores were predictors of student scores on the MEAP (McGlinchey & Hixson, 2004). The current study determined that R-CBM scores were strong predictors of MAP scores. Therefore, this study agrees with the findings of McGlinchey and Hixson (2004) that R-CBM scores are predictors of summative assessments. However, the sample for the first research question in the current study included only third grade students.

Hintze et al. (2002) correlated R-CBM, SES, gender, and white/non-white variables with the Woodcock Johnson Psychoeducational Battery-Revised (WJ-R). Hintze et al. (2002) found that the R-CBM was a strong predictor of the WJ-R for grades 2-5, but that gender, race, and SES predicted very little in regards to reading comprehension. The results of the current study mirrored Hintze et al.'s (2002) findings that R-CBM scores were the strongest predictors of MAP scores with demographic variables being weak predictors of MAP scores. Students with full pay lunch status were more likely to score higher on the MAP than students with free and reduced lunch status, in all grade levels and both subject areas except 5th grade math.

Hintze and Silberglitt (2005) researched whether R-CBM could predict scores of third grade students on Minnesota Comprehensive Assessment (MCA). Their findings indicated that the R-CBM scores of first, second, and third graders could longitudinally predict student scores on the MCA (administered at the end of third grade) (Hintze & Silberglitt, 2005). The current study supports that R-CBM scores are an accurate means of predicting student achievement on high stakes assessments in the third grade. Fall, winter, and spring R-CBM scores were strong predictors of third grade MAP communication arts assessment scores (administered in the spring).

Ax (2004) found that the R-CBM predicted passing scores on nine different state assessments. The results of the current study extended their research results by finding that R-CBM scores were strong predictors of student success on the MAP in grade 3.

The current study adds to the literature in the area of gender differences. Hardegree (2012) found that gender differences did not cause statistically significant variances in scores on the Georgia Criterion Referenced Competency Test. The current

study found that for third grade students, females were more likely to score higher on the communication arts MAP assessment than males.

Research questions two and three focused on Maze scores and demographic variables as predictors of MAP scores for fourth and fifth grades, respectively. Merino and Beckman (2010) found that the Maze assessment did not provide a statistically significant prediction for Nebraska's state assessment. The current study contradicts Merino and Beckman (2010), finding that Maze scores are strong predictors of MAP scores for fourth and fifth grade students.

Silberglitt et al. (2006) also reviewed the relationship between a summative assessment, the Minnesota Comprehensive Assessments-Reading (MCA-R), and the R-CBM and Maze assessments. A significant relationship was found between the MCA-R, the R-CBM, and the Maze assessment (Silberglitt et al., 2006). The findings of the current study are in agreement with the findings of Silberglitt et al. (2006). R-CBM and Maze scores were strong predictors of the MAP communication arts assessment scores.

Research questions four, five, and six questioned whether M-CBM scores were predictors of MAP math assessment scores. Keller-Margulis et al. (2008) and Shapiro (2012) both compared the M-CBM to the summative assessment in Pennsylvania. Their studies found a positive correlation between the M-CBM and PSSA. The findings of the current study were consistent with the findings of Keller-Margulis et al. (2008) and Shapiro (2012).

Each of the six research questions addressed in this study included demographic variables (gender, SES, and race). Literature relating to race indicated that African American and other minority students (excluding Asian students) underperform their

Caucasian counterparts (Abedi & Dietel, 2004; Hacker, 2001; Parker, 2001; Reardon & Galindo 2009). The current study determined that race was only a strong predictor of communication arts MAP scores for third grade students, in which white students were predicted to score higher than non-white students.

Conclusions

As stated in chapter one, summative and formative assessments are necessary to monitor student academic progress. This study's focus included formative and summative assessments of student achievement in both communication arts and mathematics. Implications for actions and recommendations for future research are included in this section based on the findings of this study.

Implications for action. NKCS provided the data for this study. The samples for research questions one, two, and three each had over 880 students (n = 886, 1055, and 1020, respectively). These questions focused on R-CBM and communication arts MAP scores. However, the samples for research questions four, five, and six each had less than 550 students (n = 450, 520, and 377, respectively). These questions focused on the M-CBM and mathematics MAP scores. The demographic data was the same for the two sets of research questions and all of the students were required to take the same assessments. It is recommended that NKCS determine why there is such a discrepancy in the number of students with viable communication arts and math CBM scores. Another interesting finding worth exploring has to do with the gender variables. Although females make up approximately 50% of the total population, approximately 66% of each of the sample groups was female. It is recommended that the district determines why the

percentages of males who qualified for the study are not closer to the actual percentage of the overall population.

The NKCS assesses students with the CBMs in order to monitor student progress on curricular goals. The MAP is utilized by the state of Missouri to monitor and report student progress at the state and national levels, as mandated by NCLB. The school district assesses students with the MAP assessment every spring to comply with the Missouri Department of Elementary and Secondary Education's assessment program. A goal of this research was to determine if a correlation existed between CBM scores and the MAP assessment. Since a positive relationship was found between the two assessments, it is recommended that NKCS continue using the CBM assessments as formative assessments. Furthermore, it is recommended that support, consultations, and technical assistance be provided for teachers with the continued use of this assessment.

Finally, the population sample of students from the NKCS included both students from full pay lunch status and free and reduced lunch status. These categories represent the students' SES. A positive relationship was found between a student's SES and MAP scores, in that students with full pay lunch status were more likely to score higher on the MAP than students with free and reduced lunch pay status. Therefore, it is recommended that the school districts of Missouri create programs to address the needs of low SES students.

Recommendations for future research. This research added to the literature related to formative and summative assessments. At the time of this study, No Child Left Behind required each state to develop an assessment to determine if students were learning the curriculum standards for their grade level (Myers, 2008). However, with the

development of the Common Core State Standards (CCSS) and their implementation (in 2014), a new comprehensive nationwide exam will be administered (Common Core Standards Initiative, 2012). When this assessment becomes a part of the federal mandates, this study could be replicated to determine if CBM scores, gender, race, and SES are predictors for this new nationwide assessment.

Additionally, a longitudinal study could be developed to monitor this group of third grade students as they move through the fourth and fifth grades. The purpose of this study would be to see if the score predictions created during the third grade year could predict student success in fourth grade and fifth grade, respectively. Thus, the study could be conducted to determine if third grade CBM scores and demographic data could predict success on a summative exam in fourth and fifth grades.

The current study included data for the elementary school population of the NKCS. The formative assessments utilized in this study are also utilized with the middle school students. The students in the middle school are also responsible for taking the state summative assessments. Therefore, a recommendation is that this study be replicated at the middle school level.

A further recommendation for future research includes the demographic variables for this study. The demographic variables were broken into two groupings (white/non-white, male/female, and full pay or free/reduced lunch status). With the exception of the gender variables, each of the other variables could be disaggregated into more categories. For example, SES could be categorized as full pay, reduced, and free status. Racial groups could be categorized into subgroups such as: Caucasian, African American, Native American, Pacific Islander, and Asian. Finally, two additional subgroups could be

added to the study: English Language Learners (ELL) and students with an Individual Education Plan (IEP).

AIMSweb no longer offers the use of M-CBM as an assessment measure. The NKCS now utilizes the Math Computation (M-COMP) CBM, which is a product of AIMSweb. With this occurrence, it is recommended that this study be duplicated using the M-COMP and the MAP or new Common Core State Standards Assessment.

Concluding remarks. Teachers utilize formative assessments, such as CBM, to guide their instruction. Summative assessments show the overall knowledge of what they have learned. The results of this study indicate that formative CBM scores are able to predict student achievement on the summative MAP assessment. An educator can utilize the CBM scores to adjust their instruction and help students to achieve higher scores on subsequent CBM assessments and predict for students to perform at the same level on the MAP. Suggestions for further research can provide educators with more information on how CBM can be used to help guide instruction in the future.

Educators are culturally responsive within their classrooms. Data was also analyzed to determine if gender, SES, or race could predict student achievement on the MAP assessment. Gender and race were two areas that did not predict student achievement, according to the results of data analyses. However, SES, along with CBM scores, had a predictive relationship to the MAP. Therefore, educators need to monitor students for economic related needs, which may arise within their classrooms. This knowledge of how culture and background influence assessments provides teachers with information to assist all learners to achieve success.

References

Abedi, J., & Dietel, R. (2004). Challenges in the No Child Left Behind Act for English language learners. *Phi Delta Kappan*, 85, 782-785.

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences.* Upper Saddle River, NY: Pearson Prentice Hall.

Altshuler, S. J., & Schmautz, T. (2006). No Hispanic student left behind: The consequences of "high stakes" testing. *Children Schools, 28*(1), 5-14.

Ardoin, S. P., Suldo, S. M., Witt, J., Aldrich, S., & McDonald, E. (2005). Accuracy of readability estimates' predictions of CBM performance. *School Psychology*, *20*(1), 1-22.

Ax, E. E. (2004). *Relationship between oral reading rate and statewide achievement test mastery for third grade students.* Retrieved from American Studies Commons. (ETD 944)

Bain, S. K., & Garlock, J. W. (1992). Cross-validation of criterion-related validity for CBM reading passages. *Diagnostique, 17,* 202-208.

Begeny, J. C., Codding, R. S., Dunn, E. K., Eckert, T. L., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, *43*(4), 247-265.

Brown-Chidsey, R., Johnson, P., Jr., & Fernstrom, R. (2005). Comparison of grade-level controlled and literature-based Maze CBM reading passages. *School Psychology Review*, *34*(3), 387-394.

Burke, K. (2010). *Balanced assessment*. Bloomington, IL: Solution Tree.

Christ, T. J., Johnson-Gros, K. N., & Hintz, J. M. (2005). An examination of alternate assessment durations when assessing multiple-skill computational fluency: The generalizability and dependability of curriculum-based outcomes within the context of educational decisions. *Psychology in the Schools*, *42*(6), 615-622.

Christ, T. J., & Schanding, G. T., Jr. (2007). Curriculum-based measures of computational skills: A comparison of group performance in novel, reward, and neutral conditions. *School Psychological Review*, *36*(1), 147-158.

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, *36*(1), 130-146.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39*(3), 447-462.

Clarke, B., & Shinn, D. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review*, *33*(2), 234-248.

Common Core State Standards Initiative. (2012). *Frequently asked questions*. Retrieved from http://www.corestandards.org/resources/frequently-asked-questions

CTB/McGraw-Hill. (2010a). *Missouri grade level assessments technical report 2010.* Monterey, CA: Author.

CTB/McGraw-Hill. (2010b). *TerraNova*. Retrieved from http://www.ctb.com/ctb.com/control/ctbProductViewAction?p=products&product id=745

CTB/McGraw-Hill. (2011). *Missouri achievement program: Guide to interpreting your results.* Monterey, CA: Author.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Deno, S. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15*(3), 358-374.

Deno, S. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36*(2), 5-11. Retrieved from http://proxyb.kclibrary.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ448554&site=ehost-live

Deno, S. (2003). Developments in curriculum-based measurements. *The Journal of Special Education, 37*(3), 184-192.

Deno, S., Anderson, A. R., Calendar, S., Lembke, E., Zorka, H., & Casey, A. (2002, March). *Developing a school-wide progress monitoring system.* Paper presented at the annual meeting of the National Association of School Psychologists, University of Minnesota.

Deno, S., Lembke, E., & Anderson, A. (2005). *Progress monitoring study group content module.* Retrieved from http://www.cehd.umn.edu/research/highlights/CBM/

Department of Elementary and Secondary Education. (2005). *Grade level expectation documents.* Retrieved from http://www.dese.mo.gov/divimprove/curriculum/GLEDocuments.html

Department of Elementary and Secondary Education. (2006). *Missouri school improvement program standards and indicators manual.* Retrieved from http://dese.mo.gov/divimprove/sia/msip/Fourth%20Cycle%20Standards%20and%20Indicators.pdf

Department of Elementary and Secondary Education. (2009). *Grade-level assessment.* Retrieved from http://dese.mo.gov/divimprove/assess/grade_level.html

Department of Elementary and Secondary Education. (2010a). *Assessment news & updates.* Retrieved from http://www.dese.mo.gov/divimprove/assess/

Department of Elementary and Secondary Education. (2010b). *Missouri school improvement program.* Retrieved from http://dese.mo.gov/divimprove/sia/msip/msip%20overview.htm

Department of Elementary and Secondary Education. (2011a). *2010-2011 Missouri school directory.* Retrieved from http://dese.mo.gov/directory/024093.html

Department of Elementary and Secondary Education. (2011b). *2011 grade level assessment sample parent/guardian letter.* Retrieved from http://dese.mo.gov/divimprove/assess/documents/asmt-gl-parent-guardian-sample-letter-2011.pdf

Department of Elementary and Secondary Education. (2011c). *Grade level assessments.* Retrieved from http://dese.mo.gov/divimprove/assess/grade_level.html

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 4*(2), 121-139.

Fore, C., III, Boon, R. T., Lawson, C., Sr., & Martin, C. (2007). Using curriculum-based measurement for formative instructional decision-making in basic mathematics skills. *Education*, *128*(2), 324-332.

Fuchs, D., & Fuchs, L. S. (1991). Curriculum-based measurements: Current applications and future directions. *Preventing School Failure, 35*(3), 6-11.

Fuchs, L. S. (2004). The past, present and future of curriculum based measurement research. *School Psychology Review, 33*(2), 188-192.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurements and evaluation on pedagogy, student achievement and student awareness of learning. *American Educational Research Journal*, *21*(2), 449-460.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluations: A meta-analysis. *Exceptional Children*, *53*, 199-208.

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading Maze task. *Exceptional Children, 58,* 436-450.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28,* 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, *22*(1), 27-48.

Graff, J., Jenkins, J., & Miglioretti, D. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children, 75*(2), 151-163.

Gunzenhauser, M. (2003). High-stakes testing and the default philosophy of education. *Theory into Practice, 42*(1), 51–58.

Hacker, H. K. (2001, October 19). Racial gap persists in standardized testing; in Missouri, black students had lowest scores in every subject and grade tested. *The St. Louis Post Dispatch*, B4.

Hale, A. D., Hawkins, R. O., Sheeley, W., Reynolds, J. R., Jenkins, S., Schmitt, A. J., Martin, D. A. (2011). An investigation of silent versus aloud reading comprehension of elementary students using Maze assessment procedures. *Psychology in the Schools, 48*(1), 4-13.

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review*, *32*, 228-240.

Hardegree, A. C. (2012). *Standards-based assessment and high stakes testing: Accuracy of standards-based grading* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Thesis database. (UMI No. 3540113)

Harris, D. N. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education, 113*, 367-390.

Hintze, J., Callahan, J. E, III, Matthews, W. J., Williams, S. A. S., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review, 31*(4), 540-554.

Hintze, J., & Christ, T. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review*, *33*, 204-217.

Hintze, J., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*(3), 372-386.

Howe, K. B., & Shinn, M. M. (2002). *Standard Reading Assessment Passages (RAPs) for general outcome measurement.* Eden Prairie, MN: Edformation.

Howell, K. W., Fox, S. L., & Morehead, M. K. (1993). *Curriculum-based evaluation: Teaching and decision making.* Pacific Grove, CA: Brooks/Cole.

International Reading Association. (1999). High-stakes assessments in reading. *The Reading Teacher*, *53*(3), 257-264.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and Maze. *Exceptional Children*, *59*(5), 421-432.

Johnson, B., & Christensen, L. (2008). *Educational research: Qualitative, quantitative, and mixed approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnosis accuracy of curriculum-based measures in reading and mathematics. *School Psychological Review*, *37*(3), 374-390.

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethical and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly, 14*(3), 327-342.

Lunenburg, F. C., & Irby, B. J. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences*. Thousand Oaks, CA: Corwin Press.

Marzano, R. (2009). *Formative assessment & standards-based grading.* Bloomington, IN: Solution Tree.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*(2), 193-203.

McKenna, M. C., & Miller, J. W. (1980). The effects of age and distractor type on Maze performance. In A. J. Moe & M. Kamil (Eds.), *Perspectives in reading research and instruction: Twenty-ninth yearbook of the National Reading Conference* (pp. 288-291). Washington, DC: National Reading Conference.

Merino, K., & Beckman, T. O. (2010). Using reading curriculum-based measurements as predictors for the Measure Academic Progress standardized test in Nebraska. *International Journal of Psychology: A Biopsychosocial Approach, 6*, 85-96.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*, 283–306.

Myers, S. (2008). High-stakes testing. *High-Stakes Testing -- Research Starters Education*, *1*, 1-12.

National Center for Response to Intervention. (2010). *Progress monitoring tools.* Retrieved from http://rti4success.org/progressMonitoringTools

North Kansas City Schools. (2009). *Core data report.* Kansas City, MO: Author.

North Kansas City Schools. (2010a). *Elementary assessment windows.* Retrieved from http://www.nkcschools.org/sites/default/files/assessment_plan_elementary.pdf

North Kansas City Schools. (2010b). *Official enrollment*. Retrieved from https://portal staff.nkcschools.org/staff/forms/index.phtml?pg=12&cat=111

Parker, L. (2001). Statewide assessment triggers urban school reform but how high the stakes for urban minorities? *Education and Urban Society*, *33*(3), 313-319.

Parker, R. E., & Hasbrouck, J. (1992). The Maze as a classroom-based reading measure: Construction methods, reliability, and validity. *Journal of Special Education*, *26*(2), 195-218.

Pearson Education, Inc. (2008). *Maze-CBM*. Retrieved from http://www.aimsweb.com/measures-2/maze-cbm/

Pearson PsychCorp. (2010a). *AIMSweb*. Retrieved from http://www.aimsweb.com

Pearson PsychCorp. (2010b). *Maze-CBM*. Retrieved from http://www.aimsweb.com/measures- 2/maze-cbm/

Pearson PsychCorp. (2010c). *Reading CBM*. Retrieved from http://www.aimsweb.com/measures-2/reading-cbm/

Pearson PsychCorp. (2010d). *Training session overview of math computation CBM*. Retrieved from http://www.aimsweb.com/uploads/pdfs/MCBM6pp.ppt.pdf

Pierce, R. L., McMaster, K. L., & Deno, S. L. (2010). The effects of using different procedures to score maze measures. *Learning Disabilities Research & Practice*, *25*(3), 151-160.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853-891.

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. *The California School Psychologist*, *14*, 23-34.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? *Phi Delta Kappan, 62*(9), 631-634.

Semas, J. H. (2001). High stakes testing. *Curriculum Administrator, 37*(8), 44-48.

Shapiro, E. S. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review, 43*(3), 295-311.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35.

Shin, J., Deno, S. L., & Epsin, C. (2000). Technical adequacy of the Maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34*(3), 164-172.

Shin, J., & Lee, H. (2007). Reading-growth estimates for elementary-school students using curriculum-based measurements. *Asia Pacific Educational Review, 8*(3), 353-363.

Shinn, M. R. (2004). *Administration and scoring of mathematics computation curriculum-based measurement (M-CBM) and math fact probes for use with AIMSweb.* Eden Prairie, MN: Edformation Inc.

Shinn, M. M., & Shinn, M. R. (2002a). *AIMSweb training workbook.* Eden Prairie, MN: Edformation, Inc.

Shinn, M. M., & Shinn, M. R. (2002b). *AIMSweb training workbook: Administration and scoring of reading maze for use and general outcome measurement.* Eden Prairie, MN: Edformation, Inc.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*(5), 527-535.

Smith, C. (2007). *Problem statements in the problem-solving, RTI process.* Kansas City, MO: North Kansas City Schools.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*(3), 407-419.

State of Missouri. (2009). *Missouri revised statutes*. Retrieved from http://www.moga.missouri.gov/statutes/C100-199/1600000518.HTM

Steckler, P. M. (2006). Focus on inclusion. *Reading and Writing Quarterly, 22,* 91-97.

Steckler, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15*, 128-134.

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

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for student learning.* Boston, MA: Pearson.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*(4), 498-513.

University Academy. (2009). *Missouri Assessment Program (MAP) overview*. Retrieved from http://www.universityacademy.org/gen/ua/Missouri_Assessment_Program_MAP_Overview_m144.html

U.S. Department of Agriculture. (2009) *Federal register.* Retrieved from http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs09-10.pdf

U.S. Department of Agriculture. (2010). *School meals.* Retrieved from http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs10-11.html

U.S. Department of Agriculture. (2011a). *National School Lunch Program.* Retrieved from http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs.htm

U.S. Department of Agriculture. (2011b). *Program fact sheet*. Retrieved from http://www.fns.usda.gov/cnd/lunch/

VanDeZande, J. (2010). *Update information regarding the Missouri Assessment Program.* Retrieved from http://dese.mo.gov/divimprove/assess/documents/10updateltr.pdf

Venet, C. (2009). *What is "DOK"? How does it apply to me?* Retrieved from http://www.dese.mo.gov/divimprove/curriculum/fa/documents/dok_article_show_me_art_cvenet.pdf

Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruff, A., & Murray, C. S. (2010). Differences in the relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessments for Effective Intervention, 35*(2), 67-77.

Webb, N. (2006). *Alignment, depth of knowledge, & change.* Retrieved from http://facstaff.wcer.wisc.edu/norrmw/MIAMI%20FLORIDA%20FINAL%20slides%2011-15-05.pdf

Wiley, H. I., & Deno, S. L. (2005). Oral reading and Maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*(4), 207-214.

Zane, T. W. (2009). Performance assessment design principles gleaned from constructivist learning theory. *TechTrends, 53*(1), 81-88.

Appendices

Appendix A: Household Size and Income Level for Free or Reduced Lunch Status

Table A1 shows the annual income levels in relation to the household size for students who may qualify for free and reduced lunch within the 48 contiguous states, District of Columbia, Guam, and territories. If the parent/guardian's annual income is less than the shown number for the household size, then the family qualifies for the service.

Table A1

Household Size and Income Level to Qualify for Free or Reduced Lunch Status 2009-2011

Household Size	Annual Income for Reduced Cost Meals	Annual Income for Free Meals
1	$20,036	$14,079
2	$26,955	$18,941
3	$33,874	$23,803
4	$40,793	$28,665
5	$47,712	$33,527
6	$54,631	$38,389
7	$61,550	$43,251
8	$68,469	$48,113
Each Additional Member	+$6,919	+$4,862

Note. Adapted from "Federal Register," by the U.S. Department of Agriculture, 2009, p. 13412.

Appendix B: Institutional Review Board Form - Baker University

BAKER UNIVERSITY
SCHOOL OF EDUCATION
GRADUATE DEPARTMENT

IRB PROTOCOL NUMBER _____
(IRB USE ONLY)

IRB REQUEST
Proposal for Research
Submitted to the Baker University Institutional Review Board

I. Research Investigator(s) (Students must list faculty sponsor first)

Department(s) **School of Education Graduate Department**

Name	Signature	
1. Susan Rogers		Major Advisor
2. Margaret Waterman		Research Analyst
3. Brett Knappe		University Committee Member
4. Dr. Lisa Friesen		External Committee Member

Principal Investigator:
Kristen Kuhlman Childers
Phone: 816-668-3986
Email: kchilder@nkcschools.org
Mailing address: 11009 N. McGee St. Kansas City, MO 64155

Faculty sponsor: Dr. Susan Rogers
Phone: 913-344-1226
Email: srogers@bakeru.edu

Expected Category of Review: _x_Exempt ___ Expedited ___Full

II: Protocol:

The Curriculum-Based Measurements as a Predictor for the Missouri Assessment Program (MAP) Assessments

Will the fact that a subject did or did not participate in a specific experiment or study be made part of any permanent record available to a supervisor, teacher, or employer? If so, explain.

Archived data will be used for this research study. No data from this study will be made part of any permanent record.

What steps will be taken to insure the confidentiality of the data?

To protect anonymity and insure confidentiality the application specialist in the NKCS will randomly assign a number to each student. All data provided to the researcher will remain confidential and will only be utilized by the researcher and the Baker advisory team.

If there are any risks involved in the study, are there any offsetting benefits that might accrue to either the subjects or society?

There are no risks involved in this research study. The benefit of the study is the contribution to the research related to CBM. There is limited research related to the Maze and Computation CBM and their predictive qualities for High Stakes Assessments. Although there is more information available for the Oral Reading Fluency, there is not a study available which relates it to the MAP.

Will any data from files or archival data be used? If so, please describe.

All data in this study is archival data. The archived data used in this study includes student's names (names will be replaced with randomly assigned numbers), gender, race, socio-economic status, and participation in the Oral Reading Fluency CBM (3rd grade), Maze CBM (4th and 5th grade), Computation CBM (3-5th grade) participation in Math Lab, middle school attended (each middle school is assigned a random letter), MAP achievement level and MAP scale score from sixth, seventh, and eighth grade.

Summary

In a sentence or two, please describe the background and purpose of the research.

The purpose of this study was to determine if there is a relationship between scores in reading and math on the Curriculum Based Measurements (CBM) and scores in reading and math on the Missouri Assessment Program (MAP) Assessments for third, fourth, and fifth grade students enrolled in the North Kansas City Schools (NKCS). An additional purpose was to determine if demographics (gender, race, and socioeconomic status) affect the relationship between the MAP and CBM assessments.

The study is being conducted in the NKCS. The participants of the study are students in grades 3-5 for the 2010-2011 school year. The findings could provide pertinent results to facilitate informed decision-making for the district and the continuing use of Curriculum-Based Measure (CBM) formative assessments.

Briefly describe each condition or manipulation to be included within the study.

There are no conditions or manipulations.

What measures or observations will be taken in the study? If any questionnaire or other instruments are used, provide a brief description and attach a copy.

The independent variables of the study are student scores in the Oral Reading Fluency CBM (3rd grade), Maze CBM (4th and 5th grade), Computation CBM (3-5th grade) and demographic data (gender, socioeconomic status, and race).

The research study includes a dependent variable of scores in the Communication Arts and Mathematics MAP assessment. No questionnaires or other instruments will be utilized in the study.

Will the subjects encounter the risk of psychological, social, physical, or legal risk? If so, please describe the nature of the risk and any measures designed to mitigate that risk.

Subjects will not encounter any psychological, social, physical, or legal risks in this study.

Will any stress to subjects be involved? If so, please describe.

Subjects will not experience stress during this study. All data collected is historical.

Will the subjects be deceived or misled in any way? If so, include an outline or script of the debriefing.

Subjects will not be deceived or misled in any way during this study.

Will there be a request for information, which subjects might consider to be personal or sensitive? If so, please include a description.

Student demographic data will be collected for this study. This information includes gender, race, and socio-economic status. This information will be gathered and names are omitted for student privacy..

Will the subjects be presented with materials, which might be considered to be offensive, threatening, or degrading? If so, please describe.

No materials will be presented to the subjects for this study.

Approximately how much time will be demanded of each subject?

No time will be demanded of the subjects.

Who will be the subjects in this study? How will they be solicited or contacted? Provide an outline or script of the information which will be provided to subjects prior to their volunteering to participate. Include a copy of any written solicitation as well as an outline of any oral solicitation.

The subjects in this study were third, fourth, and fifth grade students in the North Kansas City Schools from Fall 2010 until Spring 2011. Students will not be solicited or contacted for this study. The North Kansas City Schools Request for Research Team will review the findings from this IRB and decide if research can be conducted.

What steps will be taken to insure that each subject's participation is voluntary? What if any inducements will be offered to the subjects for their participation?

No subjects will be contacted for this study.

How will you insure that the subjects give their consent prior to participating? Will a written consent form be used? If so, include the form. If not, explain why not.

As no subjects will be contacted in this study, written consent is not necessary.

Will any aspect of the data be made a part of any permanent record that can be identified with the subject? If so, please explain the necessity.

All data utilized in the study will be coded for anonymity. The data analyzed for this study is already a part of the district permanent record. However, the results from the study will not be added to the permanent record.

Appendix C: Data Collection Request Form - North Kansas City Schools

Request to Conduct Research
2011-2012

Name of Applicant: Kristen Kuhlman Childers

Employee of North Kansas City Schools? Yes x No ___

If yes, location and position Maplewood Elementary School – Kindergarten Teacher

Is the research in fulfillment of graduate program requirements and/or in partnership with an external organization (e.g., university, college, business, industry, agency, etc.)? Yes x No ___

If yes, name of external organization and lead contact person:

External organization: Baker University

Lead Contact Person and Position: Dr. Susan Rogers -- *Associate Professor & Ed.D. Program Coordinator*

Purpose of research: Dissertation Topic: The Curriculum-Based Measurements as a Predictor for the Missouri Assessment Program (MAP) Assessments within grades 3-5

Submission Requirements

1. A copy of the complete application submitted for formal approval by a human subjects review board. This application should include, at a minimum:
   a. A brief summary of the purpose and scope of the research including:
   ____ The extent to which the research addresses and/or aligns with the goals of the school district
   ____ Potential benefit of the research to positively impact district, building, or classroom practice
   b. A brief summary of the research methods including:
   ____ Participants
   ____ Selection process
   ____ Remuneration procedures (if applicable)
   ____ Assurance of confidentiality of participant identification
   ____ Consent and assent procedures and documents
   ____ Activities related to the research, including proposed survey, interview, and/or assessment questions/instruments
   ____ Extent of intrusiveness/disruption regarding classroom instruction
   ____ Time/effort requirements of participants
2. _____ Evidence to demonstrate that the proposed research has been formally approved through a human subjects review process.
3. _____ Assurance from the researcher that building principals, teachers, students and/or their parents may opt out of participation without consequence even with approval by the district team.
4. _____ Assurance from the researcher that results will be communicated back to the district upon completion of study. (**Anticipated date of completion**: _____)

Date received by Director of Research, Evaluation & Accountability _____

Team Review Date: _____ Approved: ____ Not Approved: ____

Signature of Associate Superintendent: _____

_____ _____

Signature of Principal(s) of building(s) impacted by research study Date

A copy of this form must be returned to NKCS Director of Research, Evaluation, and Accountability with all necessary signatures before beginning research.

Submission Requirements as listed on the Request to Conduct Research Form

1. A. A brief summary of the purpose and scope of the research including:

- **How does the research address/align with the goals of the school district?**

The North Kansas City Schools (NKCS) has a commitment to mastery teaching and learning. To meet this goal, NKCS teachers utilize formative and summative assessments to monitor student growth. Formative assessments are designed as tools to guide instruction throughout the school year. Curriculum-Based Measurements (CBM) are a source formative assessment data which teachers utilize within NKCS. In contrast, summative assessments inform teachers of student growth at the school end of the school year. The state of Missouri requires summative assessments as part of the Missouri Assessment Program (MAP). The purpose of this study is to determine the relationship between scores on CBM and the Missouri Assessment Program (MAP) assessments for grades 3-5.
Secondly, the analysis of demographic group data (gender, race, and socio-economic status) could lead to the discovery of new information concerning student achievement within NKCS. Different subgroups may perform differently on assessments. Therefore, it is important to determine how different subgroups perform on assessments in order to find one that is the best choice for assessing all subgroups. This information is important to the culturally responsive teaching goal of the NKCS.

- **What is the potential benefit for the district to participate in this study?**

When research is complete, the findings of the research can be used to validate the usage of CBM within elementary classrooms. In addition, information

can be shared with other researchers to further the database of information surrounding formative assessments and their predictive nature towards high stakes testing. Finally, demographic data analysis will ensure that the formative assessments utilized assess all subgroups equitably.

1. B. A brief summary of the research methods including:

- **Participants**

The participants for this study are 3-5th grade students of the North Kansas City Schools during the 2010-2011 school year.

- **Selection Process**

Each student included in the population of the study must have completed three assessments. In grade 3, these assessments are the Oral Reading Fluency CBM (R-CBM), the Mathematics CBM (M-CBM), and the MAP assessment. In grades 4-5, these assessments are the Maze CBM, the M-CBM, and the MAP assessment. If a student did not complete all three assessments for their grade level, then they were omitted from the study.

- **Remuneration procedures**

None necessary for this study

- **Assurance of confidentiality of participation identification**

In order to assure confidentiality of the participants, each student will be coded by NKC employees. Therefore, the researcher did not come in contact with personal information for any of the participants.

- **Consent and assent procedures and documents**

District level consent is needed. No consent is needed from the participants.

- **Activities related to the research, including proposed survey, interview, and/or assessment questions/instruments**

The necessary information are student data scores for the following instruments: R-CBM (grade 3), Maze CBM (grades 4-5), M-CBM (grades 3-5), and the MAP assessment (grades 3-5). In addition, demographic data is required for the 3-5 grade students in the areas of race, gender, and socioeconomic status.

- **Extent of intrusiveness/disruption regarding classroom instruction**

No classroom time is required for this study.

- **Time/effort requirements of participants**

No direct contact takes place with participants.

2. Evidence to demonstrate that the proposed research has been formally approved through a human subjects review board.

- **See Attached**

3. Regarding Principal, Teacher, and Student Participation:

The research which Kristen Childers is proposing to conduct will need district level assessment data. This information will be gathered with the assistance of the Director of Research, Evaluation, and Accountability. Due to the nature of the study, all data needed is historical and coded for anonymity. Therefore, principal, teacher, and student requests for permission are not needed.

4. A. How will results of the study be communicated to the North Kansas City Schools (NKCS)?

Following the research, results will be communicated to the district in two ways.

1. **The researcher will present a copy of the findings to the Director of Research, Evaluation, and Accountability.**
2. **Dr. Lisa Friesen, NKCS District Instructional Coordinator, will be a part of the dissertation committee where the research and background will be presented.**

4. B. What is the anticipated completion date for the research?

The anticipated completion date is August, 2012.

Appendix D: Letter of Research Approval - Baker University

BAKER UNIVERSITY
Own Confidence

January 24, 2012

Kristen Kuhlman Childers
11009 N. Mc Gee St.
Kansas City, MO 64155

Dear Ms. Childers:

The Baker University IRB has reviewed your research project application (M-0124-0117-0124-G) and approved this project under Exempt Review. As described, the project complies with all the requirements and policies established by the University for protection of human subjects in research. Unless renewed, approval lapses one year after approval date.

The Baker University IRB requires that your consent form must include the date of approval and expiration date (one year from today). Please be aware of the following:

1. At designated intervals (usually annually) until the project is completed, a Project Status Report must be returned to the IRB.
2. Any significant change in the research protocol as described should be reviewed by this Committee prior to altering the project.
3. Notify the OIR about any new investigators not named in original application.
4. Any injury to a subject because of the research procedure must be reported to the IRB Chair or representative immediately.
5. When signed consent documents are required, the primary investigator must retain the signed consent documents for at least three years past completion of the research activity. If you use a signed consent form, provide a copy of the consent form to subjects at the time of consent.
6. If this is a funded project, keep a copy of this approval letter with your proposal/grant file.

Please inform Office of Institutional Research (OIR) or myself when this project is terminated. As noted above, you must also provide OIR with an annual status report and receive approval for maintaining your status. If your project receives funding which requests an annual update approval, you must request this from the IRB one month prior to the annual update. Thanks for your cooperation. If you have any questions, please contact me.

Sincerely,

Carolyn Doolittle, EdD
Chair, Baker University IRB

Appendix E: Research Approval - North Kansas City Schools

Request to Conduct Research
2011-2012

Name of Applicant: Kristen Kuhlman Childers

Employee of North Kansas City Schools? Yes x No ___

If yes, location and position Maplewood Elementary School – Kindergarten Teacher

Is the research in fulfillment of graduate program requirements and/or in partnership with an external organization (e.g., university, college, business, industry, agency, etc.)? Yes x No ___

If yes, name of external organization and lead contact person:

External organization: Baker University

Lead Contact Person and Position: Dr. Susan Rogers -- Associate Professor & Ed.D. Program Coordinator

Purpose of research: Dissertation Topic: The Curriculum-Based Measurements as a Predictor for the Missouri Assessment Program (MAP) Assessments within grades 3-5

Submission Requirements

1. A copy of the complete application submitted for formal approval by a human subjects review board. This application should include, at a minimum:
   a. A brief summary of the purpose and scope of the research including:
      ____ The extent to which the research addresses and/or aligns with the goals of the school district
      ____ Potential benefit of the research to positively impact district, building, or classroom practice
   b. A brief summary of the research methods including:
      ____ Participants
      ____ Selection process
      ____ Remuneration procedures (if applicable)
      ____ Assurance of confidentiality of participant identification
      ____ Consent and assent procedures and documents
      ____ Activities related to the research, including proposed survey, interview, and/or assessment questions/instruments
      ____ Extent of intrusiveness/disruption regarding classroom instruction
      ____ Time/effort requirements of participants
2. ____ Evidence to demonstrate that the proposed research has been formally approved through a human subjects review process.
3. ____ Assurance from the researcher that building principals, teachers, students and/or their parents may opt out of participation without consequence even with approval by the district team.
4. ____ Assurance from the researcher that results will be communicated back to the district upon completion of study. (**Anticipated date of completion**: August, 2012.)

Date received by Director of Research, Evaluation & Accountability _____

Team Review Date: 2-21-12 Approved: ✓ Not Approved: ____

Signature of Associate Superintendent: Paul Tuggen

_____ _____

Signature of Principal(s) of building(s) impacted by research study Date

A copy of this form must be returned to NKCS Director of Research, Evaluation, and Accountability with all necessary signatures before beginning research.