1-29-2015

Criterion Validity of Mathematics Curriculum-Based Measurement

Giancarlo A. Anselmo
Indiana University of Pennsylvania

CRITERION VALIDITY OF MATHEMATICS CURRICULUM-BASED MEASUREMENT

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Giancarlo A. Anselmo

Indiana University of Pennsylvania

December 2014

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Educational and School Psychology

We hereby approve the dissertation of

Giancarlo A. Anselmo

Candidate for the degree of Doctor of Education

Joseph F. Kovaleski, D.Ed.
Professor of Educational and
School Psychology, Advisor

Mark J. Staszkiewicz, Ed.D.
Professor of Educational and
School Psychology

Lynanne Black, Ph.D.
Associate Professor of Educational
and School Psychology

Timothy J. Runge, Ph.D.
Associate Professor of
Educational and School Psychology

ACCEPTED

Timothy P. Mack, Ph.D.
Dean
School of Graduate Studies and Research

Title: Criterion Validity of Mathematics Curriculum-Based Measurement

Author: Giancarlo A. Anselmo

Dissertation Chair: Dr. Joseph F. Kovaleski

Dissertation Committee Members: Dr. Mark J. Staszkiewicz
Dr. Lynanne Black
Dr. Timothy J. Runge

School districts across the United States are implementing or starting to implement Multi-Tiered Systems of Support (MTSS) to address both academics and behavior. Within this framework, valid instruments are needed to monitor both academic and behavioral progress. Curriculum-based measurement (CBM) can be a very powerful tool to universally screen students and provide a way to monitor academic progress. CBMs are generally short, fluency-based mini-assessments that provide formative data for educational planning and progress monitoring of instruction. While research has been completed in the area of reading CBM, there is a lack of evidence to support the use of mathematics CBM (M-CBM), especially at the secondary level. Research is needed to validate these instruments so they can be used for universal screening and progress monitoring (Hosp, Hosp, & Howell, 2007). M-CBM is one type of assessment that could help teachers improve monitoring of mathematics instruction and could also provide a way to identify students in need of further intervention (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006).

In an effort to obtain data on the criterion validity of secondary M-CBM, screening data were compared to the North Carolina End-of-Grade Mathematics Test (NC EOG Math Test). Correlation coefficients were completed between AIMSweb Math Calculation CBM (M-COMP) and Math Concept and Applications M-CBM (M-CAP). M-COMP and M-CAP were both taken by students at the end of the 2011 – 2012 school year. Students also took the NC EOG Math Test during the same time frame. Correlations were also completed between the M-COMP and M-CAP taken in 2011 – 2012 to the NC EOG Math Test taken at the end of the 2012 – 2013 school year.

Results of the study suggest that M-CAP has strong concurrent and predictive validity when the dependent measure is the NC EOG Math Test. The study also provides evidence that calculation skills, while important, do not have strong predictive strength, at the secondary level, when a state math assessment is the criterion. Finally, contrary to the hypothesis, concurrent validity coefficients are not higher than predictive coefficients.

Implications related to the field of school psychology and recommendations for further research are discussed

ACKNOWLEDGEMENTS

I would like to thank everyone who has made this dissertation possible. I need to start by thanking my family. To my wife Jennifer, thank you so much for your support and encouragement over this 7 year process. I can still remember sitting on the couch as we discussed the possibility of beginning this adventure. You have sacrificed so much for this to be possible, including being a single mom for months in the summers, time proofreading papers, and dealing with my emotional ups and downs as part of this process. Thank you will never be enough for your unwavering support. To my son Anthony, I have seen you grow during this process, and I am sorry for all the time I had to work when I would have rather been with you. Thank you for being so understanding of my time away. This dissertation is a testament to the fact that you can do just about anything if you set your mind to it. I love you more than I can express in words.

Thank you to all my committee members. Many of you barely knew me when I asked you to be on my committee, yet you agreed and have helped guide me through the process. Special thanks go to Dr. Kovaleski who has stuck with me through what have been countless revisions. You have

pushed me to become a better writer than I thought possible. You are a huge reason why I applied to IUP, and while this has been the hardest process of my life, I am happy that you were the one to guide me through to the end.

I also need to take time to thank my parents. Dad, thank you for instilling in me from a very young age, the importance of education and giving me first-hand knowledge about what it takes to be a master teacher. Mom, thank you for teaching me how to laugh because without this ability, I would have never made it through this process and many others throughout my life.

Finally I would like to take the time to thank my friends and colleagues. Many of you have been a great cheering section through my years in the program. To Erica, Nikole, and Samantha, I can’t tell you how happy I am to have met you three. Your support, laughter, and all important PA information have made this a more bearable and enjoyable process. To Kelsey, thank you for taking time away from your summer to help me with writing in a more professional manner. To everyone at Cleveland County Schools, thank you for giving me the opportunity over the past decade to do what I love.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Many school districts across the United States have begun implementing a Multi-Tiered System of Supports (MTSS) to address both academic and behavior concerns. Academic supports have been given within a structure that has become known as Response to Intervention (RtI). RtI is an emerging approach to structuring general education (Fuchs, Fuchs, & Compton, 2012; Reschly, 2008). In an MTSS model, a student with academic delays or behavioral concerns is given one or more research-validated interventions. Within this model, each child is targeted with evidence-based teaching practices in the regular education classroom in addition to receiving supplemental instruction outside the classroom. The student's academic and/or behavioral progress is monitored often for information regarding whether specific strategies are increasing academic or behavioral skills (Tilly, Reschly, & Grimes, 1999). If the student is not responding positively to specific interventions, the data collected are used to support an increase in the level of academic or behavioral supports the student is receiving within the school (Fuchs et al., 2012; Individuals with Disabilities Act, 2004)

RtI often involves a problem-solving component. The problem-solving model applies self-correcting processes through establishing an intervention based on scientifically-based research that is matched to student needs, implementing the intervention with fidelity, and monitoring progress. Depending on the results of the intervention, the intervention could be modified or changed if progress toward established goals is insufficient. Implementation of problem-solving within a MTSS framework is completed through the creation of a multi-tiered system that integrates general, remedial, and special education into a single system of educational instruction (Reschly, 2008). The MTSS is often depicted as a triangle. The base of the triangle would encompass the total school population and refer to evidence-based instructional practices for everyone, often including a universal screening component (McCook, 2006). The second tier, or the middle of the triangle, would serve students who based on multiple data sources, need instructional practices beyond the normal core curriculum (Fuchs, Mack, Morgan, & Young, 2003). Generally estimates suggest that 15 percent of a school population will need this level of support (Sugai, 2009). The tip of the triangle would represent the smallest number of students including those children who need the most intensive instruction. This level of support may involve special education services and formalized evaluation for special education services (Shinn, 2008). Figure 1 shows a visual representation of the triangle and how services become more intense for fewer students as you move up the triangle.

For an RtI system to work properly, it must be grounded in evidence-based instructional practices. Two components of this framework that have been identified as critical to a successful RtI system are problem-solving and curriculum-based measurements (CBMs; Tilly et al., 1999).

A problem-solving model addresses not only the needs of individual students but groups of students to address both academic and behavioral deficits. Depending on the results of the intervention, the intervention could be modified or changed if progress toward established goals is insufficient. Progress monitoring and goal setting within academics can be accomplished using CBM.

CBM is a group of standardized tests, often referred to as probes, that usually last between 1 and 5 minutes (Shinn, 2008). CBM is generally scored by counting the number of correct items on the probe in the time allotted (Deno, 2003). For example, in math, a child might have two minutes to complete several subtraction problems. The administrator counts the number of correct digits within the two minutes to obtain the score.

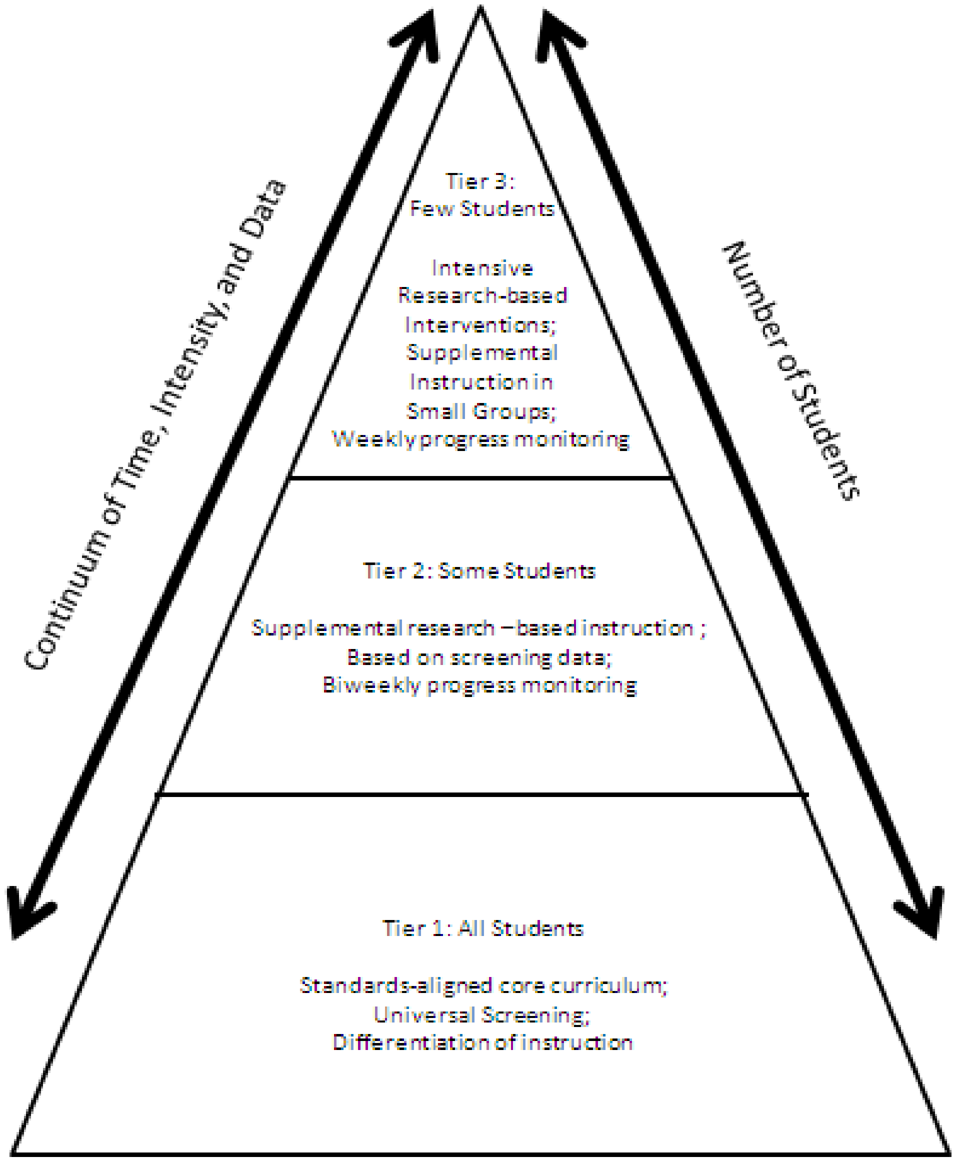

Figure 1. Three-tier model of instruction and support. Adapted from "The RtI Approach to Evaluating Learning Disabilities," by Kovaleski, J. F., VanDerHeyden, A. M., & Shapiro, E. S., 2013, p. 24. Copyright 2013 by The Guilford Press. Adapted with permission.

CBM relies on several distinguishing features. First, CBM can be used to evaluate a student's progress toward established academic goals. CBM is generally described as a general outcome measure as opposed to a summative assessment that obtains information about mastery of concepts (Stecker, Fuchs, & Fuchs, 2005). This suggests that CBM gives general data about how a student is progressing in the curriculum as opposed to measuring specific content knowledge. The second important feature of CBM is the ability to administer it often as a progress monitoring tool and the ability to graphically display the monitoring data as a way to show the response to instruction (Stecker et al., 2005). A third feature of CBM is research supporting both its reliability and validity. Using psychometrically sound instruments is imperative when educators are using the data to make educational decisions. Research has supported the use of CBM as a universal screening measure, progress monitoring tool, and for use as a survey-level assessment to drill down to specific academic deficits (Fuchs et al., 2012).

Most of the current research in regards to CBM has overwhelmingly focused on basic literacy and basic math. Within mathematics, CBM probes, generally referred to as M-

CBM, are mainly based on computation skills and designed for students prior to sixth grade. Many studies completed contain data from elementary students; these studies have consistently shown that computation probes in first through fifth grades have moderate to high reliability and validity (Clarke & Shinn, 2004; Fuchs, Hamlett, & Fuchs, 1999; Germann, & Deno, 1983; Hintze, Christ, & Methe, 2006; Jitendra, Sczesniak, & Deatlin-Buchman, 2005; Keller-Margulis, Shapiro, & Hintze, 2008; Thurber, Shinn, & Smolkowski, 2002; Tindal, Shinn & Marston, 1985).

Statement of the Problem

There is a lack of evidence to support the use of M-CBM at the secondary level for universal screening and progress monitoring (Hosp et al., 2007). M-CBM is one assessment tool that could help teachers improve monitoring of mathematics instruction and could also provide a better way to identify students in need of specialized instruction (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006). Previous research has focused on computational fluency using M-CBM designed to sample broadly from multiple computational skills within the curriculum (Helwig, Anderson, & Tindol, 2002). Studies have been completed at the secondary level with CBM focusing on mathematics.

However, the current literature base can be expanded to show that computation and application based M-CBMs are both reliable and valid measures that can be used for universal screening, monitoring progress toward goals, and for educational decision making.

The purpose of this study was to evaluate the validity of specific secondary M-CBM probes. Comparisons have been made between M-CBM and the North Carolina End-of-Grade Mathematics Test (NC EOG Math Test). The NC EOG tests are North Carolina's annual statewide assessment to obtain data concerning student achievement over the course of the academic year. If secondary M-CBM is found to be a strong predictor of success on the NC EOG Math Test, it would have broad practice implications. This finding gives the secondary schools a way to screen students early who are struggling with mathematics and at-risk of failing high-stakes tests. The CBMs could also help districts across the country with implementing RtI procedures at the middle school level. M-CBM probes need to have adequate psychometric properties so they can be used for universal screening and progress monitoring specific math instruction and interventions. M-CBM could then be an integral part of the way schools make accurate determinations about the specific skill deficits that students possess so that instruction can then be tailored to meet the specific needs of a student or group of students.

Studies completed at the elementary level support the technical adequacy of CBM (Stecker et al., 2005). Studies at this level provide evidence for strong reliability (inter-rater) and the ability of CBM to predict performance on high stakes tests (Deno, Fuchs, Marston, & Shinn, 2001). While the utility of M-CBM at the secondary level has been studied, more studies are needed to support the validity of these measures at this level (Helwig et al., 2002).

The National Council of Teachers of Mathematics (NCTM), a leader in mathematics policy, states that the two basic categories of mathematics include mathematical reasoning and specific math content. According to NCTM (2000), mathematical reasoning skills include problem-solving, communication, reasoning, and connections. Mathematical content includes number sense, computation skills, and spatial sense among others. Hudson and Miller (2006) contend that the goal of math instruction is conceptual understanding, which generally suggests both an understanding of mathematical content in addition to reasoning skills. Because both content and reasoning skills are important in learning mathematics, two types of M-CBM have been utilized in this study. The first measures math content knowledge by assessing how fluently a student can produce correct answers to grade-level math computation problems. The second M-CBM is designed to assess math reasoning skills by measuring how fluently a child can read and answer a math problem.

Comparisons have been made between M-CBM data and NC EOG Math Test data from 2012 and 2013 to obtain criterion validity for the M-CBM probes. The criterion validity data include both concurrent and predictive validity data. M-CBM data from the spring of 2012 have been correlated with NC EOG Math Test data from the end of the 2011-2012 school year. The same M-CBM data from the spring of 2012 has also been correlated with the NC EOG Math Test data from the end of the 2012-2013 school year. These comparisons establish concurrent and predictive validity data for the secondary M-CBM.

Research Questions and Hypotheses

Research Question 1

The first research question is: What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation is not predicted to be strong. Studies completed at the elementary level support moderate to strong correlations between calculation M-CBM and nationally normed math instruments/state-wide assessments (Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002). The current hypothesis contrasts with the literature because the studies reviewed do not suggest computation skills can significantly predict performance on a state-wide math assessment at the secondary level.

Research Question 2

The second research question is: What is the concurrent validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and strong correlation would exist. This hypothesis is consistent with current research that has been completed at both the elementary and secondary levels (Fuchs et al., 1994; Foegen & Deno, 2001; Fuchs et al., 1999; Helwig et al., 2002; Jitendra et al., 2005; Keller-Margulis et al., 2008; Thurber et al., 2002).

Research Question 3

The third research question is: What is the predictive validity of math computation M-CBM with the NC

EOG Math Test? Consistent with research question 1, it was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong. As stated previously this hypothesis was in contrast to studies which have been completed at the elementary level. Previous studies suggest moderate to strong correlations between computation M-CBM and nationally normed/statewide math assessments (Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002).

Research Question 4

The fourth research question is: What is the predictive validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and moderate to strong correlation would exist. This hypothesis was consistent with current criterion validity studies that have been completed at both the elementary and secondary levels (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002).

Research Question 5

The fifth research question is: Is the concurrent validity of math calculation M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. This hypothesis was based on logic suggesting that tests taken closer in time will correlate higher than ones taken at significantly different times.

Research Question 6

The sixth research question is: Is the concurrent validity of math applications M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. As stated in research question 5, this hypothesis was based on logic suggesting that tests taken closer in time will correlate higher than ones taken at significantly different times.

Problem Significance

Mathematical ability is currently seen as essential for individuals to effectively complete tasks faced in everyday life (Reyna & Brainerd, 2007). Research suggests that deficits in math start early in elementary school and then persist into adulthood. The National Assessment of Educational Progress (NAEP) assessed mathematics achievement in a nationally representative sample of 168,000 fourth-grade students and 161,000 eighth-grade students. Findings indicated that in 2009, 66% of eighth graders and 61% of fourth graders are below proficiency standards in mathematics (National Center for Education Statistics, 2009).

Overall positive life outcomes are associated proficient mathematical skills, thus increasing mathematical knowledge should be a goal for lawmakers and educators. (Reyna & Brainerd, 2007). Problems with mathematical literacy in the United States is a leading cause for losing ground to other countries in terms of science and technology expertise. Math ability is also an important aspect of overall health literacy (Reyna & Brainerd, 2007). According to the National Assessment of Adult Literacy in 2006, more than 30% of people living in the United States do not have the skills or knowledge base to make informed decisions about their health (Kutner, Greenberg, Jin, & Paulsen, 2006).

NCLB mandated that sound instructional practices are in place for both reading and math. Even though math and reading are both emphasized in curricula, the two subject areas have received significantly different resources and attention in the past. For instance, Grimm (2008) reported that the federal Reading First initiative received $6 billion in funding, but a program to expand mathematics instruction received only $1 billion in funding. Other researchers have brought attention to the lack of research that examines mathematics assessment and instruction/intervention (Fuchs, Fuchs, & Hollenbeck, 2007). This evidence supports the need for continued research on statewide testing programs, especially in mathematics, and the need to further study the instruments schools use for screening and progress monitoring mathematics instruction.

Definition of Terms

Math Computation M-CBM

Math computation M-CBMs are generally math assessments that have computation problems to be completed in a specific amount of time ranging from two to four minutes. Computation M-CBM might have one skill such as addition or have several skills (addition, subtraction, multiplication) within one probe. Specific items on a M-CBM computation probes are determined by grade specific skills that are consistent with a student's grade level curriculum. (Stecker et al., 2005).

Math Application M-CBM

Math Concept and Applications M-CBM include a selection of problems that contain multistep procedures that students need in order to incorporate addition, subtraction, multiplication, or division knowledge. Items represent grade level math curriculum with tasks that require reading items, analysis of graphs/charts, geometry, counting, applied computation, measurement, problem solving, and multi-step mathematical problems (Fuchs et al., 1994).

North Carolina End-of-Grade Mathematics Test

The NC EOG Math test is a statewide summative assessment that students take in grades three through eight. The assessment is given in the spring of each year and determines a student's level of proficiency in the area of mathematics. The 2012 NC EOG Math Test measures the skills outlined for mathematics instruction as part of the North Carolina Mathematics Standard Course of Study. The 2013 NC EOG Math Test is aligned to the Common Core Curriculum (North Carolina Department of Public Instruction, 2012). The competency goals and skills of the mathematics curriculum are divided into five separate domains at each grade level: (a) number and operations, (b) measurement, (c) geometry, (d) data analysis and probability, and (e) algebra. Scaled scores are developed from a raw score which is the number of correct answers on the test. Each grade level test has a scaled score that corresponds to the student's raw score. Scaled scores are reported with four achievement levels; these levels are predetermined and give information as to whether the student passed or failed the assessment (Bazemore, 2008).

While the standards for the NC EOG Math Test have changed, The North Carolina Department of Public Instruction reports that the two tests are very similar in design. Scale scores are still reported for both tests and students still take the test in a paper/pencil format. The tests are different in that learning objectives have changed slightly from the 2012 to 2013 school year.

Validity

The validity of a test is how well the test measures what it was designed to measure. It gives information about what can be inferred from assessment or test scores (Anastasi & Urbina, 1997). Messick (1999) states, "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment" (p. 5).

Criterion Validity

Criterion-related validity indicates how well a test can estimate or predict performance on another set of skills or tests (Anatasi & Urbina, 1997). Criterion validation of tests consists of both concurrent and predictive validity. Concurrent validity is a measure of agreement between the results obtained by one test and the results obtained for the same population on another test. Concurrent Validity is established by correlating two tests taken within the same time frame. Predictive validity is the ability of a test to predict performance on another measure or test. Anastasi & Urbina (1997) state, "The logical distinction between predictive and concurrent validation is based not on time but on the objectives of testing. Concurrent validation is relevant to tests employed for diagnosis of existing status, rather than prediction of future outcomes" (p. 119).

Multi-Tiered System of Support (MTSS)

MTSS is a system in which students are given sound academic and behavioral instruction and problems within these areas are identified and intervened upon early

(Kovaleski, VanDerHeyden, & Shapiro 2013). Different school systems have a different number of tiers while the most popular models suggest a four-or three-tiered model. Academic supports are frequently provided through a process that is known as RtI. Behavioral supports are frequently addressed under programs entitled Positive Behavior Intervention and Supports (PBIS; Sugai, Horner, & Gresham, 2002)

Response to Intervention

RtI refers to a school improvement paradigm that employs a multi-tiered service delivery model utilizing formative assessments to adjust core and supplemental instruction to ensure positive outcomes for students (Tilly, 2006). RtI is a system of assessment, curriculum, and instruction. (Kovaleski et al., 2013). Kovaleski et al. (2013) stated:

> …an RTI model would contain the following components: multiple tiers of increasingly intense, evidence-based interventions, standard-aligned core curricula, research-based, differentiated instructional strategies in general education, universal screening of students' academic skills, team-based analysis of student data using the problem-solving method,

continuous monitoring of student performance, and monitoring of treatment integrity for instruction and intervention. (p. 8)

Universal Screening

Universal screening is a systematic assessment of all students in a school on academic or behavioral skills (Ikeda, Neessen, & Witt, 2008). For the current study universal screening will reference academic skills. Assessments are generally given three times a year: at the beginning, middle and end of the school year. Universal screening is most effective when it relates to important academic attainments, is predictive of future performance, provides reliable scores, and is brief/efficiently administered (Kovaleski et al., 2013).

Progress Monitoring

Progress monitoring is a systematic procedure for tracking the progress of an academic or behavioral goal. When students are supplied with specific interventions, their progress is monitored systemically based on established academic or behavioral goals. Progress monitoring when done frequently is able to give specific information about student growth and whether students are closing the achievement gap between themselves and peers

(Tilly, Reschly, & Grimes, 1999). CBM is a research-validated way to monitor academic progress of a student or group of students.

Assumptions

This study will take into account several assumptions. The first assumption is that the assessments were administered and scored according to standardized procedures. Similarly, it is assumed that the students provided their best effort when completing all assessments. As a review of archival data, steps to ensure accurate assessment procedures were not possible. Finally, it is assumed the data will be correctly entered into the database that is used for analysis.

Limitations

Limitations of the current study relate directly to internal threats to validity. Selection is the major threat to internal validity because the current data set is one of convenience. However, the data set will be able to answer the current research questions as it contains an adequate sample size.

The predictive nature of the study also lends to further specific limitations. Multiple treatment interference could be a problem because students over the course of the year will have received different instruction based on identified weaknesses. Maturation is also problem for the study considering students have matured an entire year between assessments and not all assessments are repeated.

Delimitations

The current study also has limitations due to possible threats to external validity. The current study could be limited in its ability to generalize to larger populations because it consists of a convenience sample. An effort was made to obtain a diverse sample, but the sample consists of mostly white children because the population at the school is 70% Caucasian. Also, as only seventh-grade students will serve as participants, potential developmental differences cannot be estimated. Because the sample is between 296 and 309 students, caution must be used with any and all interpretations. Finally, generalization could be problematic given that different states have varying levels of difficulty across their state tests (Kingsbury, Olson, Cronin, Hauser, & Houser, 2003).

Summary

Discussion within this chapter contained information about how districts across the country are implementing or starting to implement a MTSS framework. Background for the use of CBM, specifically M-CBM, within this framework was discussed. Information was provided about the lack of data on M-CBM at the secondary level and the importance of increasing math performance as a whole. Finally, research questions, definitions of terms, assumptions, limitations, and delimitations were identified and discussed.

CHAPTER II

REVIEW OF RELATED LITERATURE

The purpose of the current study is to assess the criterion validity of mathematics curriculum-based measurement (M-CBM). M-CBM data can be used to help with general instruction and aid in educational initiatives, such as Multi-Tiered Systems of Support (MTSS), which is a broader umbrella term that includes both Response to Intervention (RtI) and Positive Behavior and Intervention Supports (PBIS) initiatives.

The following chapter provides background information on all aspects of the current study, starting with change in federal law and how this led to different educational initiatives. MTSS, problem-solving models, and curriculum-based measurements (CBM) are defined as they pertain to the current study including research regarding the effectiveness of CBM and the reliability and validity of these measures.

Federal Education Initiatives and Accountability

Changes in federal law have prompted educators to take a more proactive role with failing students. Since 2001 federal laws have changed the landscape of accountability within public schools across the country. For instance, the No Child Left Behind Act (No Child Left Behind [NCLB], 2002) was passed as the most current reauthorization of the Elementary and Secondary Education Act (ESEA; P.L. 89-10), which was first enacted in 1965 as a part of the War on Poverty. One of the main focuses of NCLB is Title I. Title I provides resources to help schools in educating children who are at risk. Title 1 funds are available to schools in which 40% or more of the student body falls below the poverty line (Braden & Shroeder, 2004).

Adequate Yearly Progress (AYP) is one of the main tenets of the NCLB legislation and requires schools to show they are progressing toward 100% of students meeting state proficiency standards on by 2014. If states want federal money, they have to comply with this and other NCLB mandates. Many consequences exist for schools that consistently fall behind AYP, including redirection of part of the local education agency's Title I funds, discussion with educational experts, students given the option to attend another school, the elimination of staff, and possible reorganization as a charter school (Dworkin, 2005).

As a result of NCLB, the majority of school districts in the United States are administering statewide assessments as a way to gather student outcome data (Sibley, Biwer, & Hesch, 2001). NCLB legislation is not specific in how states are expected to gather data on student performance, thus providing flexibility in the ways AYP is measured (Chubb, 2005). A portion of states combine the results of high-stakes tests across several years, while others analyze AYP from single year test scores (Dworkin, 2005). These discrepancies between states validate the need for continued investigation of the efficacy of statewide accountability models in school districts across the country. Kelley (2008) states, "inconsistencies in math state standards, curricular focus, instructional delivery, and assessment practices are reasons for large numbers of students not demonstrating expected performance outcomes" (p. 419).

NCLB and the way it changed accountability has influenced other federal legislation. Specifically with the reauthorization of the Individuals with Disabilities Education Act (IDEA, 2004), policymakers worked to align special education legislation with the mandatory accountability standards of NCLB (Braden & Tayrose, 2008). This alignment made it mandatory for states to ensure that students with disabilities are given access to instruction within the general education setting, are held to the same standards as non-disabled students, and are included in educational accountability efforts (IDEA, 2004). The language in IDEA (2004) and NCLB are similar. Both stress the use of evidence-based interventions and instruction, in addition to the provision that schools will supply better core curriculum programs for academics and behavior that stress evidence-based teaching practices for all students, thus requiring less of a need for special education services (Batsche et al., 2005)

Further changes in IDEA (2004) focus on early intervening services. Early intervening refers to identifying problems early in school and intervening so students learning can be accelerated early in their school careers. This includes services for young children and a provision that states up to 15% of their IDEA funds can be spent on prevention services for students who are found at risk for problems with academics or behavior (Batsche et al., 2005).

Multi-Tiered Systems of Support

MTSS is a general term to describe the framework in many schools use to address both academic and behavioral problems. Different systems have a different number of tiers or layers of support, while the most popular models suggest a four-or three-tiered model. The most common is a three-tiered model of service delivery which has been used to describe the restructured system in which students' academic and behavioral problems are identified and intervened upon early (Sugai, Horner & Gresham, 2002). Academic supports generally funnel through a process that came to be known as RtI, while Behavioral support frameworks go through a process known as PBIS. The RTI and PBIS approaches are generally both conceptualized as providing behavioral or academic supports along a continuum. This continuum provides good instruction for everyone and students who need more supports move up the tiers and are provided intervention at a level appropriate for their specific deficits. Combining these two initiatives under the same umbrella only makes sense because then as school systems move to implementation of MTSS they are not only considering the academic supports needed but also the social emotional supports needed for a given population of students. These components represent the foundation of a MTSS model (Higgins, 2011).

The MTSS is often depicted as a triangle. The base of the triangle encompasses the total school population and refers to evidence-based instructional practices for everyone, often including a universal screening component (Batsche et al., 2005). Screening data can be utilized to help with group problem solving or individual student problem solving. An example of a screening system that can structure data in these ways is AIMSweb (Shinn, 2008). Universal data collection can benefit instruction as students in need of both early academic or behavioral intervention are identified very early in the year. This allows schools and teachers a way to plan their resources according to specific student needs based on the screening data. Interventions at this level, whether they focus on behavior or academics are focused on whole-group instructional practices, with the intention to bring large groups of students to acceptable levels of performance (Kovaleski et al., 2013).

The second tier or the middle of the triangle targets individual students and their specific behavioral or academic difficulties. This tier generally serves students who, based on multiple data sources, need instructional practices beyond the normal core curriculum (Fuchs et al., 2003). These students generally have a group of teachers and professionals staffing their needs and making decisions about specific interventions and instructional procedures to lift them to acceptable achievement levels (Ikeda et al, 2002).

The third tier can be described as the most intensive. Kovaleski et al, (2013) states “Tier III of a multi-tier RtI system is reserved for those students who fail to make sufficient progress in Tier II. Frequently, these students need more intense interventions for longer periods of time” (p. 36). Tier III would contain the smallest number of students and include those children who need the most intensive instruction available at that school (Shinn, 2008). Instruction at this level will likely include intensive interventions, and services may include those of Title 1 and district remediation programs (Batsche et al., 2005). Intensive problem-solving is involved at this level, in addition to customized interventions based on a more extensive analysis of specific strengths and weaknesses (Kovaleski et al., 2013)

Problem-Solving

MTSS includes a problem-solving component at all levels. The process is similar at all levels but the focus changes based on the level of problem-solving. For instance, problem-solving at the Tier I level is group-based problem-solving looking at an entire grade level’s data. Moving up through the tiers, the focus narrows to specific intervention group progress (i.e., Tier II) until Tier III where the process becomes very individualized focusing on one struggling student. The problem-solving model applies self-correcting processes through establishing an intervention based on scientifically-based research that is matched to student needs, implementing the intervention with fidelity, and monitoring progress. Depending on the results of the intervention, the intervention could be modified or changed if progress toward established goals is insufficient. Implementation of a problem-solving model within this framework is completed by integrating general, remedial, and special education into a single system of educational instruction (Reschly, 2008).

Within a problem-solving model, each child is targeted with evidence-based teaching practices in the regular education classroom and research-based interventions outside of the classroom. The student's academic/behavioral progress is monitored frequently for information regarding whether specific strategies are increasing academic or behavioral skills (Tilly, Reschly, &

Grimes, 1999). If the student is not responding positively to specific interventions, the data collected are used to support an increase in the level of academic or behavioral supports the student is receiving within the school (Fuchs et al., 2012).

Curriculum-Based Measurement

Since the early 1980s, school psychologists and other educators have been searching for ways to not only find children who were struggling in content areas, but also find more sensitive ways to monitor intervention effectiveness. Cusumano (2007) reported students not meeting expected educational performance "must be identified early; at a point before the gap between expected outcomes and observed skills broadens and data must be used to identify why their learning trajectories are not progressing in the desired directions" (p. 24). Educators have sought alternatives to summative achievement tests because they cannot be given often to monitor progress, it is difficult to see small skill attainment, and they can be very time consuming to administer. (Salvia, Ysseldyke, & Bolt, 2007). According to Shinn (2008), "as schools move away from traditional systems of determining placement and services to systems with a problem-solving or solution focused orientation, the use of measurement procedures that can be administered efficiently and linked directly to intervention are required" (p. 245).

Because of the problems with summative assessments mentioned prior, schools are increasingly using CBM as part of a MTSS system designed to assess all students and identify students in need of instructional support beyond the core curriculum (Fuchs et al., 2012). While developing ways for special education teachers to monitor instruction, Deno (2003), at the University of Minnesota's Institute for Research on Learning Disability, developed CBM in the mid-1970s. CBMs are generally short, fluency-based mini-assessments that provide formative data for educational planning and progress monitoring of instruction.

CBM has been identified as a critical component to a functional problem-solving model/process (Tilly et al., 1999). CBM has several features that make it distinct from other assessment instruments. First, CBM can be used frequently to monitor progress toward both short- and long-term goals. CBM measures goals in reference to general outcomes as opposed to being a summative assessment (Stecker et al., 2005). Fuchs and Deno (1991) referred to these measures as general outcome measures (GOM) because they are fluency based and take between 1 and 10 minutes to administer.

Because of the sound psychometrics and ease of use of CBM, it has become best practice in many schools to use CBM as part of a benchmark screening process often called universal screening (Shinn, 2008). In this process, all students are tested three times per year on a wide array of basic skill measurements, which could consist of reading skills, math skills, spelling, and writing. This process allows CBM to be used proactively for the purposes of identifying students who are at risk for academic failure or have significant educational need (Fuchs et al., 2012).

Another important feature of CBM is the ability to use it often for progress monitoring which can be graphed to provide a visual representation of student progress. Stecker et al. (2005) stated:

> CBM is used in a predictive fashion to estimate whether students are on target toward meeting long-term goals; however, data also are used to judge relative current performance and to determine whether the most recent instructional program has been effective in bringing about student growth. This judgment is particularly important in special

> education because CBM data help teachers to plan and individualize instruction (i.e., tailor instruction to meet individual student's needs). (p. 797)

For instance, students who are found to be struggling in a certain area based on universal screening data can be identified, supplemental interventions can begin, and progress can be monitored frequently using CBM appropriate to the skill that is being taught.

Finally, CBM has research to support both its reliability and validity. Any assessment that is used in educational decision making should have sound psychometric properties. Research indicates that CBM can be used efficiently for universal screening, progress monitoring, and decisions regarding intervention/instructional effectiveness (Clarke & Shinn, 2004; Fuchs et al., 1999; Fuchs et al., 2000; Thompson, Roberts, Kupek, & Stecker, 1994; Thurber et al., 2002; Tindal et al., 1983). CBM measures have been developed at the elementary-level for multiple academic areas, such as basic reading skills, oral reading fluency, written expression, spelling, math computation, and math concept and applications (Stecker et al., 2005). More recent research on CBM has been done at the secondary level (Busch & Espin, 2003).

CBM is a valuable tool in the RtI process because it can be used as part of universal screening process; create school, district, and national norms; measure student achievement in relation to curriculum; and reliably track the progress of academic goals (Jewell & Malecki, 2005). Merrell, Ervin, and Gimpel (2006) observed of CBM:

> These tools have demonstrated efficacy for direct assessment and monitoring of student academic performance within the curriculum. They provide an alternative to traditional norm-referenced assessment practices and have the advantage of being more closely tied to the curriculum, they are of shorter duration, they are sensitive to incremental changes, and they can be used repeatedly to monitor growth formatively. (p. 147)

CBM measures are formative assessments. High-stakes tests are summative; often only supplying data once a year and thus, they do not provide any guidance regarding whether to change instructional practices. Because CBM is formative and a general outcome measure, it has the ability to monitor progress throughout the school year allowing for adjustments to instruction. Researchers have also determined that between CBM and teacher reports, CBM is more objective and accurate in determining which students are at risk (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006).

Since its inception CBM has been used to help teachers make decisions about the effectiveness of their instruction. However, further development and research have pushed CBM into different areas of education that it was not originally designed (Deno, 2003). CBM measures are now used effectively for universal screenings within schools to help identify not only struggling students but gaps in the curriculum (Ikeda et al., 2008). Deno (2003) states, "Recently, research has explored the use of CBM data to predict success on high-stakes assessment and to measure growth in content areas in secondary school programs and in early childhood special education" (p. 184).

Currently, most CBM research has been focused on basic skills associated with reading, writing, and math in kindergarten to sixth grade:(Foegen et al., 2007). Within mathematics, M-CBM has been mainly been designed to measure basic number sense and computation skills for use with students in preschool through fifth grades. These measures have utility in higher grades, but information is lacking on how well computation fluency relates to high-stakes state testing in mathematics, especially at the secondary level (Foegen, et al., 2005).

Mathematics Curriculum-Based Measures

The National Council of Teachers of Mathematics (NCTM), a leader in mathematics policy, stated that the two basic categories of mathematics include mathematical reasoning and specific math content. According to NCTM (2000), mathematical reasoning skills include problem-solving, communication, reasoning, and connections. Mathematical content includes number sense, computation skills, and spatial sense. Hudson and Miller (2006) argued that the goal of math instruction is conceptual understanding, which suggests an understanding of mathematical content in addition to reasoning skills. Most research of mathematics supports a multiple-factor model of mathematical assessment. Research supports the two factors as the ability to perform math facts (computation) and the ability to use math reasoning skills to complete application/word problems (Foegen et al., 2007; Thurber et al., 2002).

These two components of math have been shown to be distinct from one another. This has been shown in validity studies that demonstrate that calculation skills typically correlate higher with other measures designed to measure calculation skills; however, calculation skills do not demonstrate the same correlations with tasks designed to measure math application skills (Helwig et al., 2002). Math Concept and Applications have further been shown to be distinct from computation ability (Thurber et al., 2002). These findings (which are reviewed more thoroughly in the construct-validity section) indicate how much these two areas of math ability are distinct but related. Therefore, the current study analyzes data that contains both computation and application M-CBM. M-CBM has traditionally involved students responding to grade appropriate computation problems on probes that range in time from 2 to 5 minutes. (Hintze et al., 2006). Further analysis is needed to understand how M-CBM probes are created especially at the secondary level. Current research demonstrates two ways to currently produce M-CBM. Foegen et al. (2007) described two ways that M-CBM probes are designed. The first, termed curriculum-sampling, measures are developed by selecting items based on the grade level mathematics curriculum. For instance, in the third grade, there would be more two-digit by two-digit addition and subtraction problems, as opposed to single-digit addition and subtraction problems. This approach has been applied to computation skills and math concepts/applications. Stecker et al. (2005) states:

> Mathematics generally is accepted as very skill specific, thus content for M-CBM tests are derived by determining the grade-level skills deemed important in the student's curriculum. The general outcome is described as proficiency across these critical grade-level skills. (p. 798)

For instance, a computation probe at the secondary level might consist of multiplication up to 3 digits by 2 digits, complex addition/subtraction, and long division with remainders. Students at this level have 4 minutes to complete as many problems as they can. Scoring is performed by counting how many correct digits the student completed in the allotted time frame (AIMSweb, 2008). Within this framework, two types of probes can be used to elicit information about math performance. The first type is a single-skill computation probe (e.g., a probe consisting of multiplication problems). The second type is a mixed computation probe consisting of a mixture of what would be considered grade-appropriate computation problems.

Much of the current literature within M-CBM utilizes Monitoring Basic Skills Progress (MBSP; Fuchs, Hamlett, & Fuchs, 1999). These probes were designed using a curriculum-sampling approach very similar to the AIMSweb M-CBM probes. Two types of MBSP mathematics measures exist: Computation and Concept and Applications. The Computation measures include 30 parallel probes available in first through sixth grades. Fuchs et al. (1999) designed the measures by choosing items that were congruent with the computation skills represented by the Tennessee state curriculum at each grade level. As stated previously M-CBM has traditionally been based on computation and number sense skills and designed for students in preschool through sixth grade. While computation skills are an important aspect of mathematics, they are not always the best predictors of overall success in mathematics (Fuchs, Fuchs, & Zumeta, 2008). Because of this fact, in 1994 Fuchs and colleagues designed MBSP to focus on math Concept and Applications. The authors were concerned that only limited portions of math achievement were being addressed by the M-CBM used at the time. The Concept and Applications measure, offered in second through sixth grades, were designed in much the same way as the computation MBSP, but with a greater emphasis on application knowledge (i.e., reading graphs, understanding numbers, and solving word problems, geometry, measurement, and problem solving).

MBSP M-CBM probes are very similar to AIMSweb M-CBM probes. AIMSweb also has a Computation (M-COMP) and Concepts & Applications (M-CAP) M-CBM. The items are generally based on the NCTM principles and standards and designed using a curriculum-sampling approach (AIMSweb, 2008). These probes assess a broad set of math domains at the secondary level, including number sense, measurement, operations, patterns/relationships, data/statistics, geometry, and algebra (AIMSweb, 2008; Alonzo & Tindal, 2011).

The second approach to developing M-CBM at the secondary level is the robust-indicators approach. Foegen, Jiban, and Deno (2007) stated:

> Researchers have sought to identify measures that represent broadly defined proficiency in mathematics. Using this approach, effective measures are not necessarily representative of a particular curriculum, but are instead characterized by the relative strength of their correlations to various overall mathematics proficiency criteria. Measures in this second

> approach attempt to parallel in mathematics the kind of “robustness” that the oral reading CBM task offers in the area of reading: not necessarily drawn from the student’s yearly curriculum, yet offering strong correlations to a host of criterion measures of overall subject area proficiency. (p. 121)

Similar to the curriculum-sampling approach the robust-indicators approach is another example of a GOM. The robust-indicators approach for math at the secondary level is in its infancy, but some studies have already shown promise for this type of approach (Helwig et al., 2002).

The curriculum-sampling and robust-indicators approaches each have strengths and weaknesses. The main advantage when using the curriculum-sampling approach is it aligns with the skills expected of a child at each grade level, which facilitates a teacher’s ability to give specific information about a student’s strengths and weaknesses within the curriculum (e.g., long division with remainders; Foegen et al., 2007). For instance, M-CBM Concept and Applications are to be administered frequently over the course of the year to monitor students’ progress within math content areas (Fuchs et al., 2008). The assessment data can be collected and analyzed within the classroom, so that teachers can monitor the effectiveness of specific instruction within a short period of time of teaching the material and then use the data for future lesson planning (Thurber et al., 2002). This strength, however, can cause problems as it measures skills that are grade specific (Helwig et al., 2002). Therefore, tracking cannot be done from year to year on any one specific math skill. Also, math curricula differs greatly especially as they pertain to concepts and how they fit into the sequence of instruction (Foegen et al., 2007).

The robust-indicators method offers other advantages and disadvantages. The first advantage is the hope that the robust-indicators method can be used to demonstrate growth over the course of several years (Foegen et al., 2007). The robust-indicators approach has also demonstrated an impressive ability to predict high-stakes testing at the secondary level (Helwig et al., 2002). Unfortunately, the authors demonstrated that their concept-grounded CBM would not be beneficial for monitoring progress over time. This is a major limitation considering that progress monitoring is a hallmark of CBM. Also, because the robust-indicators are more broad based in the skills that are measured, it is difficult for educators to gain specific information about strengths and weaknesses within the current mathematics curriculum (Foegen et al., 2007).

M-CBM also has several other characteristics that lend to its utility. Using M-CBM, it is possible to identify students who are at risk in math from kindergarten to secondary levels (Clarke & Shinn, 2004). M-CBM also can effectively measure differences at different ages and grades in respect to specific math calculation skills (Hintze et al., 2006). Some research suggests that M-CBM can not only identify specific math deficits but it can do so for diverse racial groups (Evans-Hampton, Skinner, Henington, Sims, & McDaniel, 2002). Finally, M-CBM has been shown to help teachers with instruction in the classroom and many teachers find M-CBM acceptable for use within the classroom setting. Students of teachers who used M-CBM measures in the classroom showed more growth over students in classrooms that did not use progress monitoring using M-CBM (Allinder & Oats, 1997).

Reliability and M-CBM

Studies have shown M-CBM has adequate inter-rater agreement, test-retest reliability, and alternate-form reliability (Clark & Shinn, 2004; Fuchs et al., 1999;

Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, & Jancek, 2003; Tindal et al., 1983; Hintze et al., 2006; Thurber et al., 2002). In general, reliability data are reported as reliability coefficients. Reliability coefficients are a special use of the correlation coefficient and range from -1 to 1 (Salvia, Ysseldyke, & Bolt, 2009). Salvia et al. state that instruments used to monitor student progress should have a coefficient of at least .70. Greater than or equal to .80 would be the minimal level needed to make screening level decisions, and greater than or equal to .90 would be needed for important individual student decisions such as special education placement.

Inter-Rater Reliability

Inter-rater agreement, in terms of M-CBM, pertains to how consistently different people score the same M-CBM probe. For instance, if a third grader took a 2-minute, timed M-CBM that covered multiplication, would two people score them the same? Hintze et al. (2006) reported a mean inter-rater agreement coefficient of .96 when they tested 20% of their total sample of 402. The study was performed using M-CBM probes from first through fifth grades, and the M-CBMs were designed by the authors. Two types of probes were designed for the study: a single skill computation probe and a multiple skill computation probe. These probes were designed using a curriculum-sampling approach selecting computation skills appropriate at each grade level.

Fuchs et al. (2003) also demonstrated inter-rater reliability with a problem solving performance assessment M-CBM designed by the authors to be given to third grade students. The authors reported a mean inter-rater agreement coefficient of .99 when they tested 20% of their total sample of 412.

Test-Retest Reliability

Test-retest reliability refers to a test's stability over time. It is a correlation coefficient between the same tests taken at two different times. Tindal et al. (1983) reported test-retest reliability of several single-skill computation M-CBMs. The authors designed four single-skill M-CBM probes in the areas of addition, subtraction, multiplication, and division. These probes were designed using a curriculum-sampling approach for students in the fifth grade. Test-retest correlations were completed after one week and coefficients were reported for each skill and ranged from .78 to .89. Clarke and Shinn

(2004) also reported test-retest reliability between .84 and .93 for probes at the first grade level. The authors studied 52 students in the areas of Missing Number, Oral Counting, Quantity Discrimination, and Number Identification. M-CBMs used in the study were designed by the authors.

Alternate-Form Reliability

Alternate-form reliability is similar to test-retest reliability in that multiple tests are given. In the case of alternate-form reliability, a similar but not identical probe would be given within the same time frame (e.g., within a week). Thurber et al. (2002) demonstrated alternate-form reliability on two M-CBM probes designed by the authors. Two probes were designed for the study, one measuring basic math facts and another measuring computation skills along with more advanced problem solving. The authors reported alternate-form reliability between .87 and .92 for their M-CBM at the fourth-grade level with a sample of 207 students.

Fuchs et al. (1999) also demonstrated alternate-form reliability for MBSP computation measures for grades 2 through 6. The authors studied the stability of test scores on 1,089 students by analyzing the correlations between two measurements on alternate-forms administered on separate days during a 1-week interval. These correlations ranged from .73 to .88 across grade levels.

Fuchs et al. (1999) also demonstrated alternate-form reliability for MBSP computation measures for grades 2-6. The authors studied the stability of test scores on 1,089 students by correlating the average of students' first and third scores with the average of their second and fourth scores. These correlations ranged from .81 to .88.

Fuchs et al. (1999) also demonstrated alternate-form reliability for MBSP probes measuring Concept and Applications. The authors computed the average for each student's scores on odd-numbered probes and computed an average for each student's scores on even-numbered probes. The authors then derived correlations between the means for the odd probes and the even probes. The correlations ranged from .94 to .97 across grades 2 through 6.

Studies on M-CBM provide evidence for reliability at the elementary and secondary levels as stated previously. Data suggest adequate inter-rater agreement, test-retest, and alternate-form reliability. Most of the research to date has been done at the elementary level; however, with a growing concern for secondary achievement, especially in the area of math, schools are starting to investigate ways they can find and help students who are at risk at the secondary level (Hosp et al., 2007).

Validity and M-CBM

The validity of a test is how well it measures the construct it is designed to measure. The validity of test, also gives information about what can be inferred from the test scores (Anastasi & Urbina, 1999). According to Messick (1990), "Since the 1950s, test validity has been broken into three or four distinct types, one of which comprises two subtypes. These are content validity, construct-validity, and criterion validity, which is composed of predictive and concurrent validity" (p. 11).

Studies on M-CBM have established adequate validity in addition to reliability. Validity studies have been performed that have established content validity, construct-validity and criterion validity (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Shinn & Marston, 1985; Thurber et al., 2002).

Content Validity

Content validity refers to how well items on a test or assessment are a good representation of what the test has been designed to measure (Brualdi, 1999). Content validity is the assessment of items on a test to make sure the content is appropriately matched to the specific area or areas measured by the test. These procedures would frequently be used with tests that are made to assess how well someone has learned a specific skill (Anastasi & Urbina, 1990). Brualdi (1999) states "Content validity is often evaluated by examining the plan and procedures used in test construction. Did the process ensure that the collection of items would represent appropriate skills?" (p. 2). At its core, content validity is assessed by demonstrating that the items on a test measure the outcome that the test was designed to measure (Messick, 1990).

Fuchs et al. (1999) established content validity for their MBSP M-CBM using a group of educators with varied experience to provide recommendations on the design of the M-CBM. The group of educators consisted of regular and special education teachers and curriculum supervisors from two different states. This group of professionals inspected the problem types for each grade level and made suggestions which were incorporated into the design of the finished M-CBM. Based on very little modification to the M-CBM, the content validity was judged to be adequate.

Construct Validity

The construct validity of a test is how close a test comes to measuring a theoretical construct or trait (Anastasi & Urbina, 1990). Braualdi (1999) states "Construct-related validity evidence refers to the extent to which the test measures the 'right' psychological constructs. Intelligence, self-esteem, and creativity are examples of such psychological traits." (p. 2). According to Messsick (1989), "Construct-validity is evaluated by investigating what qualities a test measures, by determining the degree to which certain explanatory concepts or constructs account for performance on the test" (p. 8). Correlation between items on a test and factor analysis are generally used to show relationships between items in at test and thus show construct validity (Brualdi, 1999). Anastasi & Urbina (1997) state:

> In order to demonstrate construct validation, test developers must show that a test correlates highly with other variables with which it should theoretically correlate (convergent validity), as well as that it does not correlate significantly with variables from which it should differ (discriminant validity). (p. 121)

For example, you would expect an M-CBM probe would correlate higher with other math tests and would not significantly correlate with a test of written expression.

Thurber et al. (2002) studied the construct-validity of M-CBM. Participants consisted of 207 students in grade 4 from four northeastern elementary schools. The authors used confirmatory factor analysis to test three theoretical models for mathematics assessments. Thurber et al. (2002) described the three models:

> A unitary model where Computation and Applications comprise a general math competence construct that M-CBM measures accurately. A two-factor model where Computation and Applications are distinct constructs and M-CBM is a measure of Computation. A two-factor model where Computation and Applications are distinct constructs and M-CBM is a measure of Applications. (p. 500)

The authors utilized several different assessments within the study. An M-CBM computation probe and basic-fact probe were designed for the study. The M-CBM computation probe consisted of basic addition, subtraction, multiplication, and division problems. It also contained more advanced strategies and algorithms. Students had five minutes to complete this M-CBM. The second M-CBM probe contained just basic-facts with addition, subtraction, multiplication, and division problems. Students had two minutes to complete this M-CBM. Data were also obtained from several standardized math assessments. Students were also assessed using the Stanford Diagnostic Mathematics Test (SDMT) and the California Achievement Test (CAT). The Computation and Applications subtests were used from both assessments. Finally, students were assessed with items measuring mathematics applications from the National Assessment of Educational Progress (NAEP). The authors also used a reading CBM measure to understand reading's unique contribution to completing a math assessment.

This study had three important findings. The first finding provided evidence for convergent and divergent validity of M-CBM. The following correlation coefficients were obtained across the three M-CBM probes with the following measures: SDMT Comp ranged from .54 to .59, SDMT App ranged from .36 to .42, CAT Comp ranged from .59 to .63, CAT App ranged from .44 to .51 and NAEP ranged from .38 to .44. Also, correlation coefficients were obtained across the two basic-fact probes with the following measures: SDMT Comp ranged from .61 to .67, SDMT App ranged from .47 to .51, CAT Comp ranged from .62 to .66, CAT App ranged from .50 to .55 and NAEP ranged from .45 to .52. Finally correlation coefficients between the two probes ranged from .80 to .83. As shown, the M-CBM probes demonstrated strong correlations with other basic fact assessments and weaker correlations with nationally normed measures. Also, the M-CBM measures demonstrated weaker correlations with assessments measuring math application skills. The second finding indicated the best model for math consisted of two-factors of mathematics assessment where Computation and Application were significantly different, but similar constructs. Finally results indicated that reading CBM was also highly correlated with both Computation and Application assessments, suggesting that reading ability should not be ignored as part of math performance.

Further construct-validity evidence for M-CBM has been demonstrated. For instance, Shinn and Marston (1985) demonstrated that M-CBM probes along with reading and writing CBM differentiated students receiving different levels of instructional support. This study consisted of 209 students in general education, Title I, and special education in fourth through sixth grades. Each student in grades four and five received a math computation M-CBM and a basic multiplication M-CBM. The computational objectives were taken from the Heath Mathematics series and were operationalized cumulatively for each grade level. Students in the fifth and sixth grades also received a basic division fact M-CBM. All M-CBMs were given with a two minute time limit and scored based on the number of digits correct. The authors then analyzed the data using analysis of variance to test for significant mean differences between groups.

Results of the study indicated significant mean differences between the general education students, Title I students, and students receiving special education services. Significant mean differences were found across grade levels and across M-CBM measures. The study shows that with brief educational measures taken from curriculum decision-making accuracy is increased significantly about which students need academic support. The findings of this study have led to more widespread use of CBM not only for identification purposes but also for progress monitoring of interventions. Educators could identify when students had increased skills enough to no longer be considered at risk.

Criterion Validity

Most of the current validity evidence for M-CBM is criterion validity. The current study is also seeking evidence for the criterion validity of secondary M-CBM measures. Criterion-related validity indicates how well a test can predict or estimate performance on another set of skills or tests (Anatasi & Urbina, 1997). According to Messick (1990), “Criterion validity is evaluated by comparing test scores with one or more external variables (called criteria) considered to provide a direct measure of the characteristic or behavior in question” (p. 11). The criterion measure can be given at about the same time a test is administered (concurrent validity) or after a specific amount of time has passed (predictive validity; Anastasi & Urbina, 1990). Concurrent validity gives an indication about how a well a score can project an individual’s current ability on the specific criterion measure. Predictive validity demonstrates how well a current test is able to predict future performance on a specific criterion measure (Messick, 1989.)

Fuchs et al. (1994) after designing Concept and Applications M-CBM sought to establish criterion validity for these probes. The authors had a sample of 140 students in second, third, and fourth grades who were administered weekly probes throughout an entire school year. Criterion validity was established by comparing the Concept and Applications M-CBM to the Normal Curve Equivalent obtained by the students on the Comprehensive Test of Basic Skills (CTBS). The CTBS was administered to the students in the spring of the academic year. Correlations ranged from .63 to .81 across grade levels. This study indicated that an M-CBM focusing on Concept and Applications had very strong validity.

Fuchs et al. (1999) went on to study the validity of their MBSP probes. Criterion validity has been established for both the MBSP Computation probes and Concept and Applications probes. Initial criterion validity was established for the Computation probes by comparing the probes to the Stanford Achievement Test (SAT). Two subtests used from the SAT consisted of the Concepts of Number subtest and the Math Computation subtest. Correlation coefficients with the Concepts of Number subtest ranged from .49 to .88 across grade levels. The correlation coefficients with the Math Computation subtest ranged from .55 to .93 across grade levels.

Criterion validity was also established for the MBSP Concept and Applications probes. Participants ranged from second to sixth grade and the total sample included 235 students. Again, these probes contain items addressing counting, number concepts, names of numbers, measurement, charts and graphs, money, fractions, applied computation, and word problems. These scores were compared to scores on the CTB/McGraw-Hill Test (1997), which was administered as part of the school district's spring testing.

The study relied on three scores from CTB/McGraw-Hill Test to establish criterion validity (i.e., Mathematics Computation, Mathematics Concept and Applications and Mathematics Total Battery score). Results indicated MBSP Concept and Applications probes had a strong relationship to the CTB/McGraw-Hill measures with correlation coefficients ranging from .66 to .81 across measures and across grades.

Keller-Margulis et al. (2008) studied the relationship between benchmark assessments (i.e., basic skills data gathered from administering CBM probes to students during the fall, winter, and spring terms within a school year) and the amount of growth within one year for reading, math computation, and math Concept and Applications CBM, and a statewide achievement test. Participants came from six elementary schools in eastern Pennsylvania and ranged from first through fifth grades. The total sample included 1,461 students in the reading group and 1,477 students in the math group. The authors used AIMSweb reading probes to measure oral reading fluency. Also used were MBSP-Math Computations probes, which consisted of a single sheet of 25 mixed operation math problems. MBSP-Math Concept and Applications probes also were administered for second through fifth grades, including 18 problems designed to assess whether students had mastered Concept and Applications skills expected for their grade level. Specifically, the Concept and Applications measure addressed counting, number concepts, names of numbers, measurement, charts and graphs, money, fractions, applied computation, and word problems. These scores were compared to scores on the Pennsylvania System of School Assessment (PSSA), the measure of accountability requirements in Pennsylvania, and also to the TerraNova Achievement Test-Second Edition, with the aim of providing evidence for the validity of CBM. Keller-Margulis et al.'s study supported the ability of CBM to be used as a predictor of later performance on both state-wide assessments and commercially available instruments.

Jitendra et al. (2005) studied the validity of M-CBM word problem-solving measures in addition to calculation probes. Participants consisted of 77 students in third grade from a northeastern state in a suburban school district. Jitendra et al. used a word problem-solving probe that consisted of addition and subtraction problems consistent with the third grade text books. Students had ten minutes to complete eight problems. Students also took a computation probe which allowed three minutes to complete 25 problems with a total of 50 correct digits possible. Criterion measures for this study consisted of mathematics problem-solving and procedure subtests from the Stanford Achievement Test-9 (SAT-9). In addition, the study also used the Mathematics Computation and Concept and Applications subtests of the TerraNova Achievement Test. Results indicated that the word problem probes showed both concurrent and predictive validity when comparing scores to both the SAT-9 and TerraNova Achievement Test (.48 - .71). Computation M-CBM also showed correlation coefficients ranging from .38 -.64. Jitendra et al.'s study supports the psychometric properties of M-CBM. It also suggests adequate criterion validity of both computation-based probes and math word problem probes.

Reliability and Validity of Secondary M-CBM

Empirical support for M-CBM has been reviewed; however, few studies have researched the effectiveness of M-CBM to the secondary level (Foegen & Deno, 2001; Helwig et al., 2002). Foegen and Deno (2001) studied the psychometric properties of M-CBM measures at the middle school level. M-CBM measures in this study examined basic-facts, basic estimation ability, and modified estimation tasks which incorporated word problems. Criterion measures for this study consisted of student grades, scores from the CAT, and teacher ratings. Results indicated that all probes had adequate reliability in respect to internal consistency, test-retest, and parallel forms. Results of the criterion validity analyses indicated that moderate relations existed between students' M-CBM scores and the measures of grades, CAT scores, and teacher ratings (.30 to .62). Basic-fact M-CBM showed correlation coefficients ranging from .44 to .63. Foegen and Deno's study provides a foundation for use of M-CBM at the secondary level.

Helwig et al. (2002) studied 171 eighth graders using a robust-indicator M-CBM probe. The authors' study took 11 items from a large sample of, concept-grounded math problems and correlated the results with a computer adaptive test. The computer adaptive test was developed with help from the Oregon Department of Education for use in the study as a substitute because the state-wide assessment results had not been completed. Correlations for regular education students (r = .80; n = 90) and students with specific learning disabilities (r = .61; n = 81) were strong to very strong. This study indicates that M-CBM can demonstrate strong criterion validity at the secondary level and shows promise for the robust-indicator method of M-CBM development.

Some work has also started to expand the use of M-CBM at the secondary level to more content specific applications within high school settings. Foegen, Olson, and Impecoven-Lind (2008) have outlined a process for designing M-CBM measures that can be used for progress monitoring algebra. This work is specific to the Project Algebra Assessment and Instruction: Meeting Standards, which is a federally funded project to design and study the effectiveness of progress monitoring measures in algebra. The authors' review provided information about development, initial growth studies, collaboration with teachers, and recommendations for future design and research.

Reliability and Validity of AIMSweb M-CBM

No peer reviewed studies could be found that document the technical adequacy of AIMSweb M-CBM probes; however, Pearson, the publisher of AIMSweb materials, has published a technical manual that outlines the specifics of test construction and pilot studies used to determine the reliability and validity of the M-CBM measures they currently publish (Pearson, 2012).

M-COMP

M-COMP M-CBM have been designed to measure computation fluency of students in first through eighth grades. The probes have been designed using a curriculum-sampling approach based on recommendations from the NCTM. Table 1 displays the content measured at each grade level. Reliability has been established for M-COMP probes based on the results of three pilot studies reported in the technical manual (Pearson, 2012). Alternate-form reliability was reported for the probes at each grade level. Coefficients range from .82 to .90 across first through eighth grades. With sample sizes ranging from 919 to 1,048 across grade levels.

Table 1

Content Measured by M-COMP, by Grade

Domain	Grade 1	2	3	4	5	6	7	8
Column addition	x	x	x					
Basic-facts	x	x	x	x	x	x		
Complex computation	x	x	x	x	x	x	x	
Decimals				x	x	x	x	x
Fractions				x	x	x	x	x
Conversions					x	x	x	x
Percentages					x	x	x	x
Integers						x	x	x
Expressions						x		
Reducing						x	x	
Equations							x	x
Exponents							x	x

Note. Adapted with permission from "Aimsweb Technical Manual," p. 18. Copyright 2012 by Pearson.

Criterion validity has also been reported for M-COMP probes in first, third, and eighth grades. During the pilot studies, the M-COMP probes were correlated with the Group Mathematics Assessment and Diagnostic Evaluation (G-MADE). At each grade level the correlations were strong; first grade had a coefficient of .84, third grade had a coefficient of .73, and eighth grade had a coefficient of .76.

M-CAP

M-CAP M-CBM have been designed to measure mathematical concept and application knowledge. These probes have been designed based on recommendations from NCTM with specific mention of the 2006 publication, *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence*. Furthermore the probes were designed based on the five math strands identified by the National Research Council. These five strands are: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. Table 2 describes the domains measured at each grade level. Three pilot studies reported in the technical manual were used to further validate the probes (Pearson, 2012). Reliability and validity were established based on national field testing. All probes were administered during field testing to a sample of 6,550 students from across the country. Inter-rater and alternate-form reliability have been reported for M-CAP. Inter-rater reliability was established by taking a random sample of 60 cases from the overall field test sample. The M-CAP probes were then scored by two administrators who had training from the M-CAP Administration and Scoring Guide. Inter-rate

Table 2

Content Measured by M-CAP, by Grade

	Grade							
Domain	1	2	3	4	5	6	7	8
Number Sense	x	x	x	x	x	x	x	x
Operations	x	x	x	x	x	x	x	x
Patterns & Relationships	x	x	x	x	x	x	x	x
Measurement	x	x	x	x	x	x	x	x
Geometry	x	x	x	x	x	x	x	x
Data & Probability	x	x	x	x	x			
Algebra					x	x	x	x
Probability						x	x	x
Data & Statistics						x	x	x

Note. Adapted with permission from "Aimsweb Technical Manual," p. 26. Copyright 2012 by Pearson.

reliability coefficients were very high ranging from .97 to .99 across grades two through eight. Alternate-form reliability was established by taking the average correlation between the probes given during field testing. Coefficients range from .81 to .88 across grade levels with a median reliability of .86.

Criterion validity was also reported in the technical manual for M-CAP (Pearson, 2012). Criterion validity was established using data from districts that use the AIMSweb database. M-CAP screening measures (fall, winter, and spring) have been correlated with the NC EOG Math Test in third grade and also with the Illinois Standards

Achievement Test (ISAT) in grades four through eight. Reported correlation coefficients with the NC EOG Math Test ranged from .60 at the beginning of the year and .64 at the middle and end of year. Results from the comparisons with the ISAT are similar. Ranges for coefficients at each grade level across the year were: fourth grade ranged from .56 to .60, fifth grade ranged from .60 to .65, sixth grade ranged from .74 to .78, seventh grade ranged from .74 to .80, and eighth grade ranged from .73 to .76.

Overall data reported in the technical manual suggests that AIMSweb M-COMP and M-CAP probes have strong reliability and validity. Data from the technical manual is consistent with previous research suggesting that M-CBM has moderate to strong reliability in the areas of computation and concepts/applications.

Summary

Over the past several decades, as educators and policymakers have become focused on accountability and outcomes for all students, many school systems are taking a second look at math instruction and progress monitoring at the secondary level (Deno, 2003). By using M-CBM, educators might be better equipped to serve the math needs of secondary students.

CBM can be used to help identify students in need of extra intervention. By incorporating CBM within MTSS models, CBM can be used proactively to screen for skill deficits and thereby facilitate early intervention to students in need. Given this emphasis, researchers should continue to investigate the effectiveness of CBM for academic screening. As indicated by Wallace, Espin, McMaster, Deno, and Foegen (2007), "The breadth and depth of CBM research varies….Substantial research has been conducted in the elementary grades; less has been conducted in the secondary grades. Reading has received more attention than has mathematics" (p. 66). Despite the lack of attention given to CBM, and to mathematics in particular, the literature offers a wide range of research to assess performance across different populations of students. These findings reinforce the literature that supports the use of CBM probes as screening devices to be used within an RtI model to inform early intervention.

Because many districts across the country are moving to implement MTSS at the middle-school level, research must be performed to support the use of certain CBM (Wallace, et al., 2007). In order for school districts to implement RtI with both integrity and fidelity, they must use assessment instruments that can not only predict performance on high-stakes tests, but also supply teachers a vehicle with which to monitor progress quickly and efficiently over the course of the year.

Much evidence and research exists for M-CBM at the elementary level (Foegen et al., 2007; Stecker et al., 2005). However, currently, little evidence exists that supports CBM use at the secondary level, especially in the area of math (Rutherford-Becker & Vanderwood, 2009). Data exists for reliability and validity of M-CBM at the secondary level, but as described in this chapter, the research is in its early stages. The current study seeks to garner evidence for M-CBM at the secondary level by demonstrating criterion validity (predictive and concurrent) to statewide assessments.

CHAPTER III

METHODS AND PROCEDURES

In the current study, the criterion validity of math curriculum-based measurements (M-CBMs) was investigated. The study includes an appraisal of both concurrent and predictive validity of M-CBM using the North Carolina End-of-Grade Test of Mathematics (NC EOG Math Test) as a dependent measure. Data from the study can be used to help with general math instruction and aid in educational initiatives such as Response to Intervention (RtI). It aids practitioners who use RtI at the secondary level by providing information about the psychometric properties of commercially available M-CBM.

The methods and procedures used to answer the research questions included in this study are described in this chapter. A description of the study site is provided including location, demographics, and instructional characteristics. Additionally, descriptions of the instruments used in this study are presented. Finally, a review of the research questions, the analyses used to answer the questions and discussion about assumptions are described.

Design

This study is a criterion validity study of M-CBM. M-CBM data measuring calculation skills (M-COMP) and math Concept and Applications (M-CAP) were analyzed to provide evidence for both concurrent and predictive validity of M-COMP and M-CAP M-CBM. Concurrent validity was studied while students were in the seventh grade and took all assessments including the NC EOG Math Test. Then predictive validity was studied when students took M-CBM measures in the seventh grade and scores were compared to the NC EOG Math Test taken in the eighth grade. Data were analyzed using Pearson Product Moment Correlations and Fisher z transformations.

Population

Archival and anonymous data from the 2011 – 2012 and 2012 - 2013 school years were examined in this study. This study represents a sample of convenience as only archival data were analyzed. The data were collected from a rural middle school consisting of sixth, seventh, and eighth grades in southwestern North Carolina. During the 2011 - 2012 school year, approximately 988 students were served by this school (National Center for Educational Statistics

[NCES], 2014). Less than 2% of students were identified as English Language Learners (ELL). Twelve percent of students in this school were serviced with an Individualized Education Program (IEP) in both the 2011 – 2012 and the 2012 – 2013 school years. Special education enrollment at this school was slightly lower than the North Carolina state average of 13.6% (North Carolina Department of Public Instruction [NCDPI], 2014). There was a 15.4 to 1 average student-to-teacher ratio during the 2011 – 2012 school year (NCES, 2012). Approximately 53.4% of students received free and reduced lunches during the 2011 – 2012 school year.

White/Non-Hispanic students comprised the majority of the student population at 72.1% (NCES, 2012). Approximately, 20.4% of the student population was made up of African-American students with 3.2% of the student population reporting to be of Hispanic descent. Less than 1% identified as an Asian/Pacific Islander.

Sample

Inclusion Criteria

Data from students who took the regular administration of the NC EOG Math Test during the 2011 - 2012 and 2012 -

2013 school years and also took the M-CBM during their seventh grade year were included in the study.

Exclusion Criteria

Students taking alternate versions of the NC EOG Math Test were excluded from the study. The North Carolina Extended Content Standards Test, also referred to as the NC EXTEND 2 has the same content as the regular test but is given in an easier format. Students who had this testing accommodation were excluded from the study because the difficulty level of EXTEND 2 is markedly different than the regular EOG.

Participants

Seventh grade has been chosen because, in most cases, this is the middle year of most middle schools. It is not the point of entry or the last year before leaving middle school. Also, most research that has been completed on M-CBM extends only to sixth grade, so this is a natural upward extension of the current body of literature (Helwig et al., 2007).

Description of Sample

Data were collected and analyzed for students who took the NC EOG Math Test and M-CBM at the end of their seventh grade year. Students' M-CBM scores taken at the end of seventh grade were then compared to the NC EOG Math Test taken at the end of eighth grade. In total, 311 students took both sets of assessments. Students who completed the M-CBM in their seventh grade year and the NC EOG Math Test during their eighth grade year totaled 298.

The following tables display the analyses of demographic data collected on students based on the year of the NC EOG Math Test. Demographic data of students were gathered prior to the current study and were part of the archival information obtained for analysis. As shown in Table 3, data on student age were reported for 311 students in 2012 and 298 students in 2013.

Table 3

Age of Participants by Year-Frequency Distributions

Age	2012		2013	
	n	%	*n*	%
12	71	22.8	0	0.0
13	208	66.6	71	23.8
14	31	10.3	198	63.7
15	1	0.3	28	9.0
16	0	0.0	1	0.3
Total	311	100.0	298	100.0

The mean age for the total sample of students in 2012 was 12.88 (SD = .572); the mean age for the total sample of students in 2013 was 13.86 (SD = .573). Ages ranged from

12 years to 15 years in 2012 and from 13 years to 16 years in 2013.

According to analysis of sex (see Table 4), the total sample from 2012 was comprised of 172 males and 139 females (55.3% and 44.7%, respectively). The total sample from 2013 was comprised of 166 males and 132 females (55.5% and 44.5%, respectively).

Table 4

Sex and Ethnicity of Participants by Year-Frequency Distributions

	2012		2013	
Sex and Ethnicity	*n*	%	*n*	%
Sex				
Male	172	55.3	166	55.5
Female	139	44.7	132	44.5
Total	311	100	298	100
Ethnicity				
White	224	72.0	214	71.6
African American	66	21.2	64	21.4
Hispanic or Latino	10	3.2	10	3.4
Multi-Racial	11	3.6	10	3.6
Total	311	100	298	100

The ethnic composition of the total sample from 2012 and 2013, as shown in Table 4, was comprised primarily of White students. The next highest number was that of African American students followed by Hispanic/Latino students and, finally, multiracial students.

Category of student disability is displayed in Table 5. In 2012, the majority of students had no disability while 19 students in the sample had a disability, making up

6.5% of the sample. In 2013, the majority of students had no disability while 18 students in the sample had a disability, making up 6.1% of the sample.

Table 5

Disability Category of Participants by Year-Frequency Distributions

Disability	2012		2013	
	n	%	*n*	%
No Disability	291	93.5	281	93.9
Learning Disability	15	4.8	15	5.0
Other Health Impairment	4	1.2	2	0.6
Speech Impaired	1	0.5	1	0.5
Total	311	100	298	100

Free/reduced lunch status is displayed in Table 6. For the total sample in 2012, the majority of students had either free or reduced lunch (*n* = 156, 50.2%) while 155 students in the sample received regular lunch, making up 49.8% of the sample.

Table 6

Free/Reduced Lunch Status of Participants by Year-Frequency Distributions

Disability	2012		2013	
	n	%	*n*	%
Regular	155	49.8	140	46.9
Reduced	16	5.2	18	6.2
Free	140	45.0	140	46.9
Total	311	100	298	100

For the total sample in 2013, the majority of students again had either free or reduced lunch (*n* = 158, 53.1%)

while 140 students in the sample received regular lunch, making up 46.9% of the sample.

Measurement

Dependent Variable

The NC EOG Math Test was used as the dependent variable. North Carolina has been testing math as part of EOG testing in grades 3 through 8 for over a decade (*North Carolina Blue Ribbon Commission on Testing and Accountability*, 2008). The 2012 NC EOG Math Test measures the skills outlined for mathematics instruction as part of the North Carolina Mathematics Standard Course of Study, while the 2013 EOG Math Test is aligned to the Common Core Curriculum (Common Core State Standards, 2010). The competency goals and skills of the mathematics curriculum are divided into five separate domains at each grade level: (a) number and operations, (b) measurement, (c) geometry, (d) data analysis and probability, and (e) algebra. Scaled scores are derived from a raw score or a number correct score for the test.

The NC EOG Math Test is generally re-normed every five years *(North Carolina Blue Ribbon Commission on Testing and Accountability*, 2008). This explains the difference in standards between years 2012 and 2013. Students receive two scores on the test, both providing different ways of reporting overall performance. First the students receive a scaled score, which ranged from 300 to 380 in 2011 – 2012 and based on the recent revision of the test, scores range from 430 to 480. Scaled scores are based on the student's raw score which is the number of correct items on the test. Each grade level has a scaled score which corresponds with each student's raw score. Scaled scores are then placed into one of four achievement levels (Bazemore, 2008). Achievement levels are reported as a 1, 2, 3 or 4. Level 3 and 4 are considered passing, and 1 and 2 are considered not passing or performing below grade level (Bazemore, 2008). The NC EOG Math Test is taken over two days and contains both calculator-active and calculator-inactive portions. These portions are combined and no breakout scores are reported. The most recent data reported for the NC EOG Math Test indicated reliability coefficients between .90 and .92 for sixth through eighth grades (Bazemore, 2008). The NCDPI reports average internal consistency (coefficient alpha) reliability of .92. Criterion-related validity has also been reported based on teacher judgments of expected grades and expected achievement levels. These scores range from .55 to .66 in sixth through eighth grades based on both teacher judgments of grades and expected levels of performance on the NC EOG Math Test (Bazemore, 2008). Data suggest that the NC EOG Math Test is a reliable and valid measure of math ability; however, the current test only gives students an overall broad score. This broad score makes it difficult to identify specific deficits the child may have; however, it is an appropriate measure to answer the current research questions.

Independent Variables

The independent variables for this study are published and designed by AIMSweb and are part of their data system used for universally screening all students within a grade level. This study analyzed the relationship between M-COMP and M-CAP taken in at the end of seventh grade (2011 – 2012 school year) to the NC EOG Math Test taken at the end of seventh grade. It also analyzed the relationship between M-COMP and M-CAP from the end of seventh grade (2011 – 2012 school year) to the NC EOG Math Test taken at the end of eighth grade (2012 – 2013 school year). AIMSweb M-CBM are intended to be used in the universal screening of all students at the beginning, middle, and end of the school year, and then used for frequent progress monitoring of students identified as having specific weaknesses (AIMSweb, 2012).

M-COMP. The M-COMP is a brief, standardized test of math operations that are part of the typical curriculum at grade 7. The seventh grade M-COMP probe will consist of several different computation skills that could be expected from a typical seventh grader. The probes will have multiplication up to three digits by two digits, complex addition/subtraction, and long division with remainders. Students have 4 minutes to complete the probe. The scoring is completed by counting how many correct digits are placed in the correct place column in 4 minutes. M-COMP measures the latent variable of calculation skills. This instrument measures how fluently a student can produce correct answers to grade level math computation problems.

The developers report alternate-form reliability coefficients of .90 for seventh grade M-COMP probes. Coefficients range from .82 to .90 across grades first through eighth (Pearson, 2012). AIMSweb also reports criterion validity coefficients for first grade at .84, third grade at .73, and eighth grade at .76. Correlations are based on comparison of the M-COMP probes to the Group

Mathematics Assessment and Diagnostic Evaluation (Pearson, 2012).

M-CAP. M-CAP probes include a selection of problems that include multi-step procedures that students need in order to incorporate addition, subtraction, multiplication, or division knowledge. These probes are based on the National Council of Teachers of Mathematics (NCTM) Principles and Standards. Seventh grade level probes include 33 problems consisting of items measuring: probability, algebra, number sense, operations, patterns/relationships, measurement, and geometry. (AIMSweb, 2009). The probes were designed to sample the math curriculum over an entire academic year; in this way all skills are represented on each probe. Students have 10 minutes to complete the probe. When scoring M-CAP, items are either correct or incorrect and each item has a point score related to difficulty (1 point for the least difficult, 2 points for medium difficulty, or 3 points for the most difficult items) based on item difficulty. This scoring process does not rely on or allow for partial-credit scoring, which significantly decreases the scoring time (AIMSweb, 2009). The authors state that the scoring procedure outlined minimizes problems with students skipping items, by providing proper weighting of the items students choose to answer.

Developers report alternate-form reliability coefficients for these probes as ranging from .80 to .88 (Pearson, 2012). Coefficient reported for seventh grade is .88. Inter-rater reliability is also reported between .97 and .99 across grade levels. Criterion validity coefficients are reported based on comparison to the Illinois Standards Achievement Test. Coefficients range from .56 to .80 in grades four through eight. Coefficients for seventh grade range from .74 to .80.

Procedure

Existing archival and anonymous data were used for this study. The data were gathered and archived by school personnel prior to and independent of this study. Therefore, no procedures were applied to subjects by the researcher.

First, the principal investigator provided the county testing coordinator with a copy of the approved project proposal and discussed the desired participant parameters and desired data. Using the databases maintained at the district office, the school district's testing coordinator created a list of students meeting the parameters of the study, added NC EOG Math Test scores, and the M-CBM data. Data were collected on students who took the M-CBM and NC EOG Math Test during the 2011 – 2012 school year when students were in seventh grade. Data were also collected on students who took the M-CBM in 2011 -2012 and also took the NC EOG Math Test during the 2012 – 2013 school year when students were in eighth grade. The testing coordinator used his/her access to district databases to compile all data into one spreadsheet. The testing coordinator identified students originally by North Carolina Window of Information on Student Education (NCWISE) number. The NCWISE number is an identification number assigned to all public school students in the state of North Carolina. The testing coordinator then removed identifying information from the data before giving the data to the principal investigator. The NCWISE number was changed to numbers starting at 1 and ranging up to the number of students in the study. Data were then sent as an email attachment from the testing coordinator to the principal investigator. The principal investigator then saved the data on his personal laptop for analysis using the Statistical Package for the Social Sciences (SPSS) software. The file sent was also coded with a password so only the principal investigator was able to open the file. At no time was the primary researcher given access to personally-identifiable information.

Data Analysis

The following section will review the research questions and hypotheses for each research question. Table 7 summarizes research questions, instruments, and variables within the study.

Research Question 1

The first research question is: What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation is not predicted to be strong. Studies at the elementary level support moderate to strong correlations between calculation M-CBM and nationally normed math instruments/state-wide assessments (Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002). The current hypothesis contrasts with the literature because the studies reviewed do not suggest computation skills can significantly predict performance on a state-wide math assessment at the secondary level. With this question the researcher studied the latent variable of math computation

Table 7

Research Questions, Latent Variable, Observed Variable, Instrument/Source, Validity and Reliability

Research Questions	Latent Variable	Observed Variables	Validity	Reliability
1. What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test?	Computation fluency	M-COMP scores	Criterion = .76	Alternative Form = .90
	Math achievement	Scale score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92
2. What is the concurrent validity of math applications M-CBM with the NC EOG Math Test?	Math Application Knowledge	M-CAP scores	Criterion = .80	Alternate Form = .88 Inter-rater = .99
	Math achievement	Scale score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92
3. What is the predictive validity of math computation M-CBM with the NC EOG Math Test?	Computation fluency	M-COMP scores	Criterion = .76	Alternative Form = .90
	Math achievement	Scaled score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92
4. What is the predictive validity of math applications M-CBM with the NC EOG Math Test?	Math Application Knowledge	M-CAP scores	Criterion = .80	Alternate Form = .88 Inter-rater = .99
	Math achievement	Scaled score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92
5. Is the concurrent validity of math calculation M-CBM different from its predictive validity?	Computation fluency	M-COMP scores	Criterion = .76	Alternative Form = .90
	Math achievement	Scaled score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92
6. Is the concurrent validity of math calculation M-CBM different from its predictive validity?	Math Application Knowledge	M-CAP scores	Criterion = .80	Alternate Form = .88 Inter-rater = .99
	Math achievement	Scaled score on the NC EOG Math Test	Criterion = .62	Coefficient Alpha = .92

knowledge in which the observed variable was M-COMP scores. The researcher was also concerned with the latent variable of math achievement which was measured by the NC EOG Math Test scale score.

Research Question 2

The second research question is: What is the concurrent validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and strong correlation would exist. This hypothesis is consistent with current research that has been completed at both the elementary and secondary levels (Foegen & Deno, 2001; Fuchs et al., 1994; Fuchs et al., 1999; Helwig et al., 2002; Jitendra et al., 2005; Keller-Margulis et al., 2008; Thurber et al., 2002).

With this question the researcher studied the latent variable of math application knowledge in which the observed variable was M-CAP scores. The researcher was also concerned with the latent variable of math achievement which was measured by the NC EOG Math Test scale score.

Research Question 3

The third research question is: What is the predictive validity of math computation M-CBM with the NC EOG Math Test? Consistent with research question 1, it was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong. As stated previously this hypothesis was in contrast to studies which have been completed at the elementary level. Previous studies suggest moderate to strong correlations between computation M-CBM and nationally normed/statewide math assessments (Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002).

With this question the researcher studied the latent variable of math computation knowledge in which the observed variable was M-COMP scores. The researcher was also concerned with the latent variable of math achievement which was measured by the NC EOG Math Test scale score.

Research Question 4

The fourth research question is: What is the predictive validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and moderate to strong correlation would exist. This hypothesis was consistent with current criterion validity studies that have been completed at both the elementary and secondary levels (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis, et al., 2008; Thurber et al., 2002).

With this question the researcher studied the latent variable of math application knowledge in which the observed variable was M-CAP scores. The researcher was also concerned with the latent variable of math achievement which was measured by the NC EOG Math Test scale score.

Research Question 5

The fifth research question is: Is the concurrent validity of math calculation M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. This hypothesis was based on logic suggesting that tests taken at a similar time will correlate higher than ones taken at significantly different times.

Research Question 6

The sixth research question is: Is the concurrent validity of math applications M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. This hypothesis is based on logic suggesting that tests taken at a similar time will correlate higher than ones taken at significantly different times.

Statistical Analysis

Following the collection and organization of data in a spreadsheet format, it was imported into the SPSS software for analysis. Initially, descriptive statistical analyses were generated for the purposes of collecting demographic information. Then four Pearson Product Moment Correlations were computed to address the first four research questions. Finally Fisher z transformations were completed to answer the last two research questions.

The purpose of correlation is to determine the strength of a relationship between two quantitative variables. This particular statistical process helped determine both the concurrent and predictive validity of the M-COMP and M-CAP probes when the NC EOG Math Test was used as a dependent measure. The fifth and sixth research questions and hypotheses were examined by Fisher z transformations. This analysis converts each correlation coefficient to a z-score. This is completed using Fisher's formula to convert a Pearson Product Moment Correlation to a z-score. Significance testing could then be obtained based on the z-score that resulted from the analysis (Preacher, 2002).

Assumptions for a Pearson Product Moment Correlation and Fisher z transformations include the use of interval or ratio data, normal distributions in each data set, minimal outliers and linear data. The first assumption is the use of interval/ratio data. The instruments used in this study produce interval data. Descriptive statistics, frequency distributions, and histograms were generated to examine the normality of the M-COMP, M-CAP, and NC EOG Math Test data. The normality of the data set was first assessed by visually inspecting the frequency distributions and histograms. Skewness and kurtosis statistics were also analyzed to determine normality. Scores for all instruments were converted to z-scores to check for outliers within the data. Linearity was checked by a visual inspection of the scatterplots of the data.

Summary

The purpose of this study was to gather information about the concurrent and predictive validity of M-CBM (M-COMP and M-CAP) probes using the North Carolina NC EOG Math Test as a dependent measure. Research questions posed relate to the ability of M-CBM probes to predict performance on the on the NC EOG Math Test. To complete this research, archival data were gathered on students who took the NC EOG Math Test during the 2011/2012 and 2012/2013 school years. These students also took the selected AIMSweb probes as part of a universal screening procedure in the spring of their seventh grade year. Student data were compiled and sent to the principal investigator with all identifying information removed so that results were analyzed anonymously. Descriptive statistics, scatterplots, and correlations were used to examine the criterion validity of both M-COMP and M-CAP. Table 8 summarizes the research questions, hypotheses, variables, and analyses used in the current study.

Table 8

Research Questions, Hypotheses, Variables, Statistical Analyses, and Statistical Assumptions

Research Questions	Hypotheses	Variables	Statistical Analyses	Statistical Assumptions
1. What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test?	M-COMP will not demonstrate a strong correlation with the NC EOG Math Test	M-COMP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2012	Pearson Correlation	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity
2. What is the concurrent validity of math applications M-CBM with the NC EOG Math Test?	M-CAP will demonstrate a moderate to strong correlation with the NC EOG Math Test	M-CAP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2012	Pearson Correlation	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity
3. What is the predictive validity of math computation M-CBM with the NC EOG Math Test?	M-COMP will not demonstrate a strong correlation with the NC EOG Math Test	M-COMP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2013	Pearson Correlation	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity
4. What is the predictive validity of math applications M-CBM with the NC EOG Math Test?	M-CAP will demonstrate a moderate to strong correlation with the NC EOG Math Test	M-COMP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2013	Pearson Correlation	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity
5. Is the concurrent validity of math calculation M-CBM different from its predictive validity	Concurrent correlation will be stronger than predictive correlation	M-COMP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2012 and 2013	Fisher z transformations	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity
6. Is the concurrent validity of application M-CBM different from its predictive validity	Concurrent correlation will be stronger than predictive correlation	M-CAP scores from spring 2012 and NC EOG Math Test Scaled Scores from spring 2012 and 2013	Fisher z transformations	1. Interval or ratio data 2. Normal distributions 3. Minimal outliers 4. Linearity

CHAPTER IV

DATA AND ANALYSIS

The North Carolina End-of-Grade Mathematics Test (NC EOG Math Test) was used as the dependent measure in assessing the concurrent and predictive validity of mathematics curriculum-based measurements (M-CBM). Specifically, research questions addressed the extent to which mathematics computation (M-COMP) and mathematics application (M-CAP) curriculum-based measurements (CBMs) predict student performance on the NC EOG Math Test. The generated correlation coefficients give specific information pertaining to the criterion validity of these M-CBM probes at the secondary level. These data will help educators determine whether such probes can and/or should be used as screening measures in the area of mathematics at this level. In this chapter, results of the investigation of the relationship between M-CBM and the NC EOG Math Test are presented.

Results of Analyses

Test of Assumptions of Statistical Procedures

Data were analyzed using Pearson Product Moment Correlations and Fisher's z transformations. Assumptions for these statistical procedures include the use of interval or ratio data, normal distributions in each data set, linearity, and minimal outliers. M-CBM and the NC EOG Math Test both produce interval data.

The normality of the data was first assessed by visual inspection the frequency distributions and histograms. See figures 2-5 for histogram data.

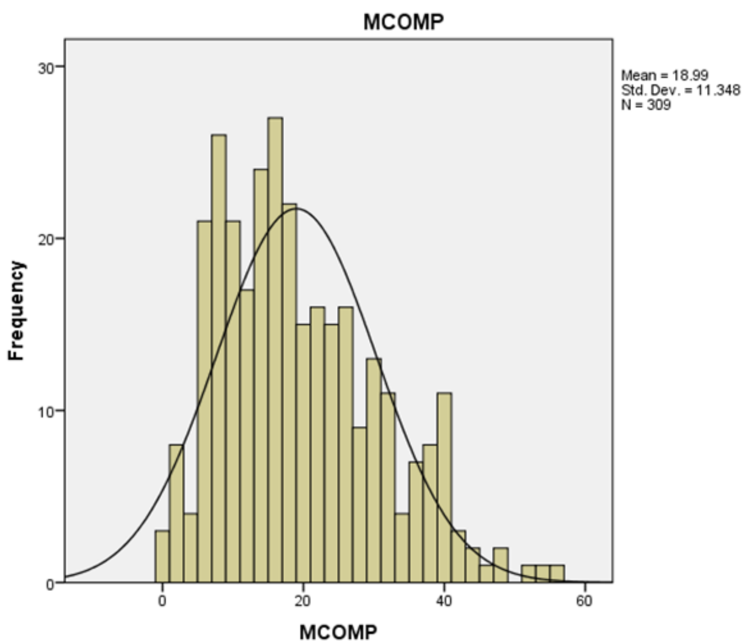

Figure 2. Histogram of M-COMP data.

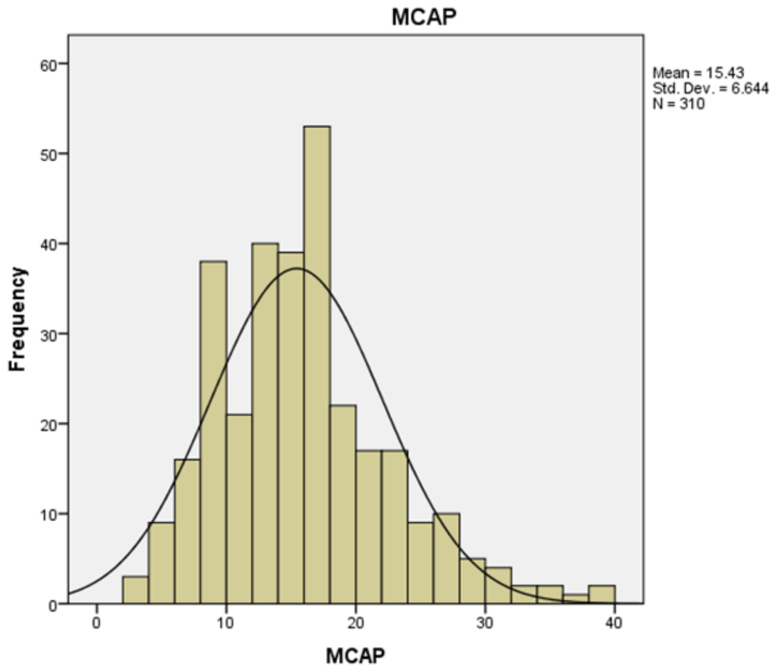

Figure 3. Histogram of M-CAP data.

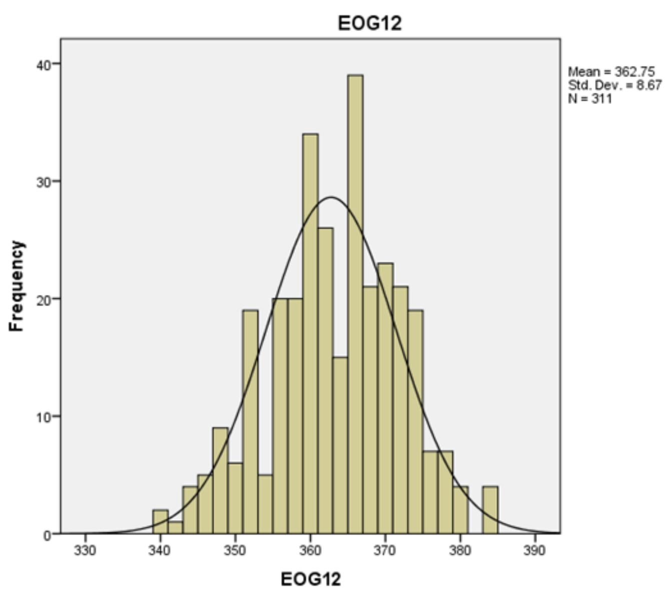

Figure 4. Histogram of NC EOG Math Test data 2012.

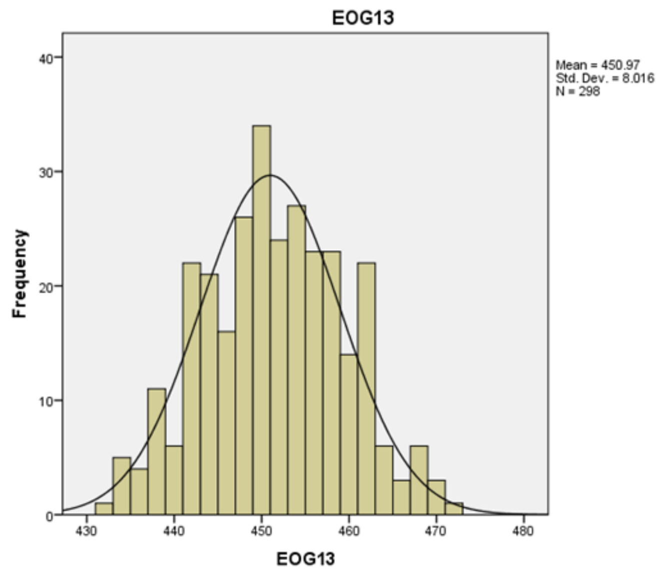

Figure 5. Histogram of NC EOG Math Test data 2013.

Each of the frequency distributions of the M-CBM data and the NC EOG Math Test scores appeared to be normally distributed based on histograms and frequency distributions. Skewness and kurtosis statistics were also run to further establish the normality of the data. The normal range for skewness has been defined as between -1 to 1 (Breakwell, 2006). As seen in Table 9 skewness statistics ranged from -.18 to .76 across measures. These scores fall within the acceptable range, indicating that the distributions were symmetrical. Kurtosis statistics were also analyzed. Kurtosis statistics between -3 to 3 is accepted as falling within the normal range (Gaur & Guar, 2006).

Table 9

Descriptive Statistics for M-COMP, M-CAP, and NC EOG Math Test

	Range	$\bar{X}$	*SD*	Skewness	Kurtosis
M-COMP	0-56	18.99	11.35	.66	-.08
M-CAP	3-39	15.43	6.64	.76	.80
NC EOG 2012	340-383	362.75	8.67	-.18	-.33
NC EOG 2013	432-471	450.97	8.01	.02	-.40

Note. M-COMP = AIMSweb math computation probe; M-CAP = AIMSweb math concepts and application probe; NC EOG 2012 = 2012 North Carolina End-of-Grade Mathematics Test; NC EOG 2013 = 2013 North Carolina End-of-Grade Mathematics test.

Across measures, values for kurtosis ranged from -.40 to .80 and fell well within the acceptable range (see Table 9). Skewness, kurtosis, histograms, and frequency distributions all indicate that M-CBM and NC EOG Math Test data were normally distributed.

The data set contained very few outliers. Outliers were analyzed by converting the raw scores for the 2012 and 2013 NC EOG Math Test, M-COMP, and M-CAP data to standardized values or z-scores. This value indicates how many deviations a score is away from the mean. For the purposes of this study an outlier was determined to be more than three standard deviations from the mean score. The NC EOG Math Tests had no scores above 2.5 standard deviations for both the 2012 and 2013 tests. The M-COMP data had two scores that were more than three standard deviations from the mean and the M-CAP data had three scores that were more than three standard deviations from the mean. These outliers remained within the data set for analysis because there were so few, and they did not significantly affect the normality of the data based on the skew and kurtosis data.

Linearity was analyzed by visual inspection of the scatterplots. Scatterplot data suggest that linear relationships exist between the dependent and independent variables. See figures 6-9 for scatterplot data. Descriptive statistics were further analyzed as shown in Table 6. Taken together, this information suggests that the data met the assumptions necessary to run Pearson Product Moment Correlations and Fisher z transformations to answer the specific research questions.

Criteria for Determining Strength of Correlations

The first four research questions data were analyzed using Pearson Product Moment Correlations. Evans (1996) has suggested the following ranges to determine the strength of correlation coefficients. Weak correlations have been defined as falling between .20 and .39, moderate correlations have been defined as falling between .40 and .59, strong correlations have been defined as falling

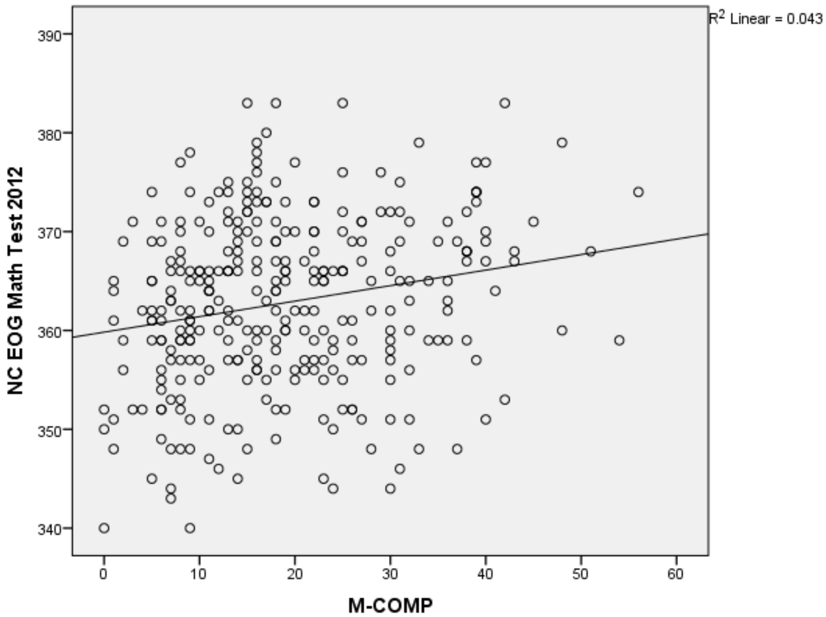

Figure 6. Scatterplot of 2012 Math Test scores and M-COMP.

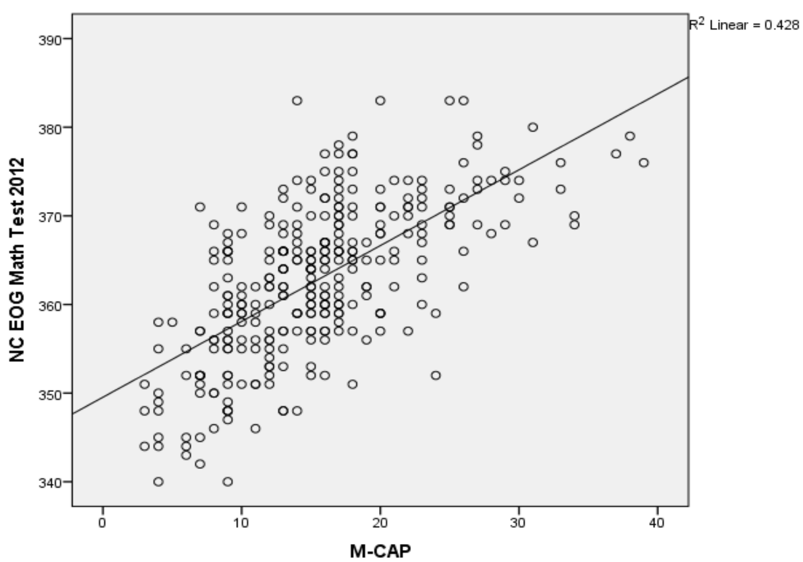

Figure 7. Scatterplot of 2012 Math Test scores and M-CAP.

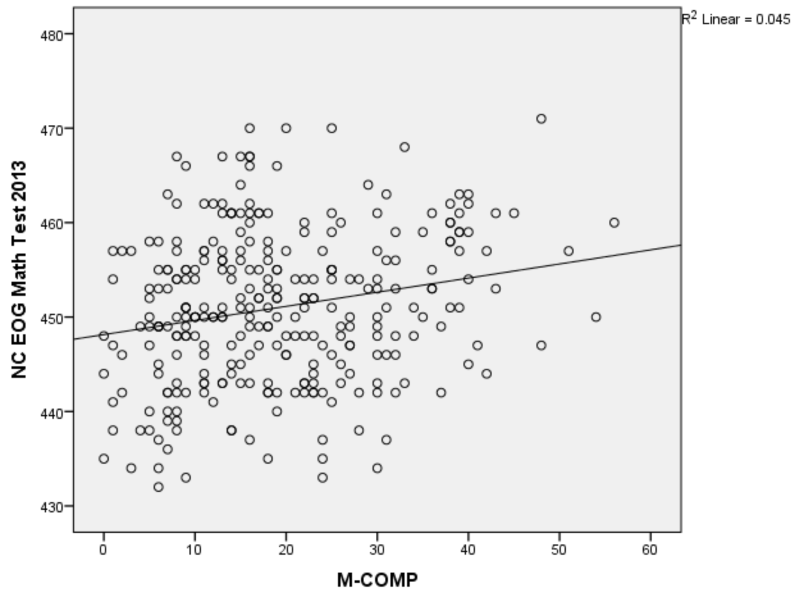

Figure 8. Scatterplot of 2013 Math Test scores and M-COMP.

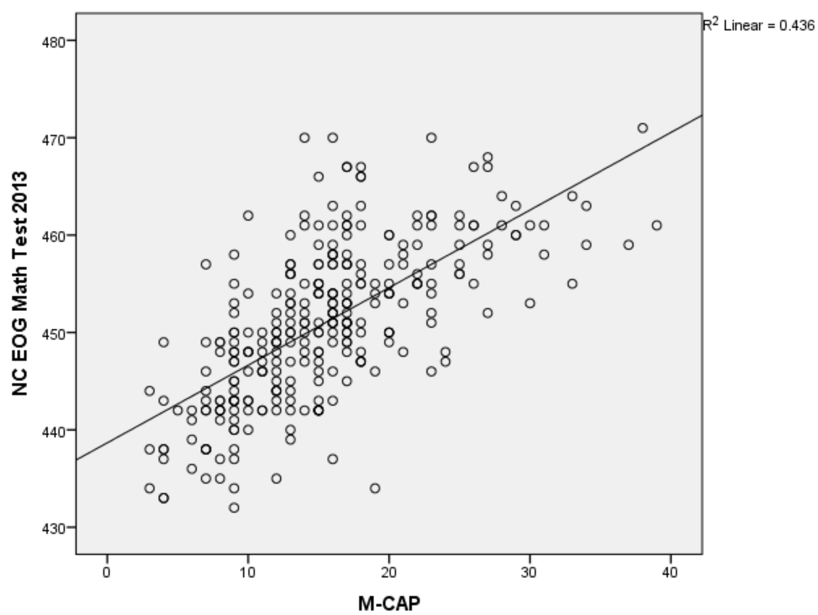

Figure 9. Scatterplot of 2013 Math Test scores and M-CAP.

between .60 and .79. Very strong correlations have been defined as above .79.

Research Question 1

The first research question was: What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong.

The M-COMP taken in seventh grade and NC EOG Math Test data for seventh grade were analyzed using the Pearson Product Moment Correlation. The correlation between M-COMP taken in seventh grade and the NC EOG Math Test (taken in 2012) was found to be statistically significant, $r(307) = .207$, $p < .01$, two tailed. M-COMP accounted for 4.3% of the variance in NC EOG Math Test scores from 2012 ($r^2 = .043$). It was hypothesized that this correlation would be statistically significant but not particularly strong. This hypothesis was supported.

Research Question 2

The second research question was: What is the concurrent validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and strong correlation would exist.

The M-CAP taken in seventh grade and NC EOG Math Test data for seventh grade were analyzed using the Pearson Product Moment Correlation. The correlation between the M-CAP from seventh grade and the NC EOG Math Test (taken in 2012) was found to be statistically significant and strong, $r(308) = .654$, $p < .01$, two tailed. M-CAP accounted for 42.6% of the variance in NC EOG Math Test scores from 2012 ($r^2 = .426$). It was hypothesized that this correlation would be statistically significant and strong. This hypothesis was supported.

Research Question 3

The third research question was: What is the predictive validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong.

The M-COMP from seventh grade and NC EOG Math Test data for eighth grade were analyzed using the Pearson Product Moment Correlation. The correlation between the M-COMP taken in seventh grade and the NC EOG Math Test (taken in 2013) was found to be statistically significant, $r(294) = .213$, $p < .01$, two tailed. M-COMP accounted for 4.5% of the variance in NC EOG Math Test scores from 2013 ($r^2 =$

.045). It was hypothesized that this correlation would be statistically significant but not particularly strong. This hypothesis was supported.

Research Question 4

The fourth research question was: What is the predictive validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and moderate to strong correlation would exist.

The M-CAP taken in seventh grade and NC EOG Math Test data for eighth grade were analyzed using the Pearson Product Moment Correlation. The correlation between the M-CAP taken in seventh grade and the NC EOG Math Test (taken in 2013) was found to be statistically significant and strong, $r(295) = .660$, $p < .01$, two tailed. M-CAP accounted for 43.6% of the variance in NC EOG Math Test scores from 2013 ($r^2 = .436$). It was hypothesized that this correlation would be statistically significant and moderate to strong. This hypothesis was supported.

Research Question 5

The fifth research question was: Is the concurrent validity of math calculation M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. This hypothesis is based on logic suggesting that tests taken closer in time will correlate higher than ones taken at significantly different times.

The difference between the concurrent and predictive validity of M-COMP were analyzed using the Fisher z transformation. The concurrent and predictive validity of M-COMP were not significantly different ($Z = -0.08$ $p > .05$, two tailed). It was hypothesized that there would be a significant difference. This hypothesis was not supported.

Research Question 6

The sixth research question was: Is the concurrent validity of math applications M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. This hypothesis is based on logic suggesting that tests taken closer in time will correlate higher than ones taken at significantly different times.

The difference between the concurrent and predictive validity of M-CAP were analyzed using the Fisher z transformation. The concurrent and predictive validity of M-CAP were not significantly different ($Z = -0.13$ $p > .05$, two tailed). It was hypothesized that there would be a significant difference. This hypothesis was not supported.

Summary

In this chapter, the analyses used to answer the research questions were discussed. First, descriptive statistics were completed to describe the sample and check for specific assumptions related to the statistical procedures needed to answer the research questions. Analysis of the data suggested the sample met the assumptions needed to analyze the data using Pearson Product Moment Correlations. The first four research questions were answered using Pearson Product Moment Correlations and Fisher z transformations were then computed to answer research questions five and six.

The correlations between M-COMP measures and the NC EOG Math Test were significant, but weak, when analyzing both concurrent and predictive validity. See Table 10 for a summary of correlation coefficients.

Table 10

Summary of Concurrent and Predictive Correlation Coefficients

	NC EOG Math Test 2012	NC EOG Math Test 2013
M-COMP	.207*	.213*
M-CAP	.654*	.660*

Note. M-COMP = AIMSweb math computation probe; M-CAP = AIMSweb math concepts and application probe; NC EOG 2012 = 2012 North Carolina End-of-Grade Mathematics Test; NC EOG 2013 = 2013 North Carolina End-of-Grade Mathematics test. **p* < .01, two-tailed.

Confirming the stated hypothesis, these correlations were significant, but not strong. The correlations between M-CAP measures and the NC EOG Math Test were significant and strong when analyzing both concurrent and predictive validity. Confirming the stated hypotheses, these correlations were significant and strong. Finally, examination of the correlation coefficients suggested virtually no difference in the concurrent and predictive validity coefficients between the 2012 and 2013 school years. These findings were in contrast to the hypotheses that stated assessments taken in closer temporal proximity would have higher correlation coefficients than those taken with a significant time lapse between test administrations.

CHAPTER V

DISCUSSION

Introduction

The current research study investigated the criterion validity of mathematics curriculum-based measurements (M-CBM) using the North Carolina End-of-Grade Mathematics Test (NC EOG Math Test) as a dependent measure. Specifically, concurrent and predictive validity of mathematics computation M-CBM (M-COMP) and mathematics application M-CBM (M-CAP) were examined. In this chapter a review of the research, findings from data analyses, discussion of results, implications for educators, limitations of the study, and suggestions for future research are presented.

Overview

Changes in federal law have prompted educators to take a more proactive role in assisting struggling learners. As a result, structures have been put into place in many schools that address both academic and behavioral concerns of its students. While such structures are known by many names, Multi-Tiered Systems of Support (MTSS) is a general term to describe such frameworks. Within the academic sector of MTSS, such structured models are often part of a system known as Response to Intervention (RtI). RtI often refers to a school improvement paradigm that employs a multi-tiered service delivery model utilizing formative assessments to adjust core and supplemental instruction to ensure positive outcomes for students (Tilly, 2008). It is a framework of instruction used to monitor student academic progress after the implementation of research-based academic or behavioral interventions (Daley, Martens, Barnett, Witt, & Olsen, 2007). It can be employed in various academic subjects including, but not limited to, reading, math, and writing.

Public schools are increasingly using curriculum-based measurements (CBMs), along with other evidenced-based tools, as part of an RtI model designed to assess the general student population and provide early intervention to children whose educational needs are beyond the scope of what the general curriculum can provide (Fuchs et al., 2012). While many of these measures have shown adequate psychometric properties, research has been lacking in the area of math CBM (M-CBM), especially at the secondary level (Foegen, Jiban, & Deno, 2007). M-CBM with sound psychometric properties will be needed as MTSS are implemented more frequently at the secondary level.

The research undertaken in this study investigated the concurrent and predictive validity of two types of M-CBM probes: AIMSweb math computation probes (M-COMP) and math application probes (M-CAP). These validity studies were completed by comparing M-COMP and M-CAP probes to the NC EOG Math Test, which served as the criterion measure. The study utilized archival data from the 2011-2012 and 2012-2013 school years.

Research Questions and Hypotheses

Research Question 1

The first research question was: What is the concurrent validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong. Analysis of this research question confirmed the stated hypothesis.

Results from this study indicate that math calculation skills in seventh grade do not correlate strongly with a high stakes math test. This finding is in contrast to previous research that has been completed at the elementary level (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis et al., 2008). The authors of previous research conducted at the elementary level have found significant and moderate to strong correlations between computation M-CBM and either statewide math assessments or nationally normed math assessments. In contrast, in this study while correlations were significant they were much weaker than previous studies.

Several possible reasons exist for this finding in the current study. Computation skills are a necessary part of mathematics; however, they are inadequate even at the elementary level (Fuchs et al., 2008). Despite this finding, early elementary school math curricula are largely based on a student's computation skills (Cawley, Parmar, Foley, Salmon, & Roy, 2001). Therefore, math assessments are mainly focused on measuring students' pure computation ability and not reasoning. Early math computation skills will more naturally be related to calculation fluency. In middle school, students are expected to apply these basic computation skills to more complex math word problems and multi-step analyses. Secondary math assessments focus more on application skills that are not necessarily represented by math calculation skills.

Another reason as to why the M-COMP probe and the NC EOG Test did not have strong correlations may be explained by the way in which questions on the NC EOG Test are presented. High stakes math tests at the secondary level are affected by students' reading abilities due to the nature of how questions must be communicated to the test-taker (Jordan et al., 2002). This suggests such assessments are influenced by a wide array of abilities including computation proficiency, reasoning skills, and scores are further affected by overall reading ability. For instance, math word problems not only require understanding of math language but also understanding the linguistics used within the problems (Levine & Reed, 1999).

Research Question 2

The second research question was: What is the concurrent validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and strong correlation would exist. Analysis of this research question, using a Pearson Product Moment correlation, confirmed the stated hypothesis. Results from this study indicate that math application skills in seventh grade did correlate strongly with a high stakes math test as a comparison between the two assessments resulted in a correlation coefficient of $r = .65$. This finding suggests evidence exists for the concurrent validity of AIMSweb M-CAP probes.

Findings from the current study are consistent with previous research (Fuchs et al., 2000; Fuchs et al., 1998; Jitendra, et al., 2005; Thurber et al., 2002). Authors of these studies have shown that application M-CBM have moderate to strong correlations with nationally normed math assessments and state-wide assessments. Current and previous findings suggest that application M-CBM appears to align well with higher-order math skills.

Research Question 3

The third research question was: What is the predictive validity of math calculation M-CBM with the NC EOG Math Test? It was hypothesized that a significant correlation would exist; however, the correlation was not predicted to be strong. Analysis of this research question confirmed the stated hypothesis.

Results from this study indicate that math calculation skills in seventh grade do not correlate strongly with a high stakes math test taken in eighth grade. This finding is in contrast to previous research that has been completed at the elementary level (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra et al., 2005; Keller-Margulis et al., 2008). The authors of previous research conducted at the elementary level have found predictive validity evidence for M-CBM when a statewide math assessment or nationally normed math assessment is the dependent measure. Keller-Margulis et al. (2008) demonstrated correlation coefficients as high as r = .69 when comparing computation M-CBM to the Pennsylvania System of School Assessment (PSSA) in third through fifth grades. In contrast, coefficients from the current study were much weaker.

Several possible reasons exist for this finding. As stated previously, computation skills are a necessary part of mathematics; however, they are inadequate even at the elementary level (Fuchs et al., 2008). Beyond elementary school, students are expected to apply these basic skills to more complex math word problems and multi-step analyses. Secondary math assessments focus more on application skills that are not adequately represented by just math calculation skills presented in isolation.

Another reason as to why the M-COMP probe taken in seventh grade and the NC EOG Test taken in eighth grade did not have a strong correlation may be explained by the way in which questions on the NC EOG Test are presented. High stakes math tests at the secondary level are affected by students' reading abilities due to the nature of how questions must be communicated to the test-taker (Jordan et al., 2002). As stated previously, secondary math assessments are influenced by a wide array of abilities including, but not limited to, computation proficiency, reasoning skills, and scores are further affected by overall reading ability.

Research Question 4

The fourth research question was: What is the predictive validity of math applications M-CBM with the NC EOG Math Test? It was hypothesized that a significant and moderate to strong correlation would exist. Analysis of this research question, using a Pearson Product Moment correlation, confirmed the stated hypothesis. Results from this study indicate that math application skills in seventh grade did correlate strongly with a high stakes math test with a resultant correlation coefficient of $r = .66$. This finding suggests evidence exists for the predictive validity of AIMSweb M-CAP probes.

Findings from the current study are consistent with previous research (Fuchs et al., 1994; Fuchs et al., 1999; Jitendra, et al., 2005; Keller-Margulis et al., 2008; Thurber et al., 2002). Authors of these studies have shown that application M-CBM has moderate to strong predictive validity when compared to either a high-stakes math assessment or norm referenced math assessment. Current and previous findings suggest that application M-CBM appears to align well with higher-order math skills.

Research Question 5

The fifth research question was: Is the concurrent validity of math calculation M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity.

Analysis of this research question did not confirm the stated hypothesis. Analysis of the correlation coefficients showed virtually no difference when comparing the correlation of the math calculation CBM taken in the seventh grade and the NC EOG Math Test taken in the seventh grade and the correlation of the math calculation CBM taken in the seventh grade and the NC EOG Math Test taken in the eighth grade. It was not anticipated that the correlation differences would be minimal given the amount of time that had passed between the administration of the M-CBM in seventh grade and the NC EOG Math Test in eighth grade.

The consistency between the concurrent and predictive correlation coefficients on both the M-CAP and M-COMP measures is noteworthy. This result may indicate that this particular school district did not use the data from M-CBM for academic planning or intervention. As these probes are intended to help assess student need in particular areas, students who scored low on them in 2011-2012 should have been given evidence-based interventions before taking the NC EOG Math Test again in 2012-2013. If this were the case, they would be expected to perform better, thus showing a less consistent correlation.

Research Question 6

The sixth research question was: Is the concurrent validity of math applications M-CBM different from its predictive validity? It was hypothesized that the concurrent validity would be significantly different from the predictive validity. Analysis of this research question did not confirm the stated hypothesis.

Consistent with calculation skills, analysis of the correlation coefficients showed virtually no difference when comparing the correlation between the math calculation CBM taken in the seventh grade and the NC EOG Math Test taken in the seventh grade and the correlation between the math calculation CBM taken in the seventh grade and the NC EOG Math Test taken in the eighth grade. Again, it was not anticipated that the correlation would be virtually identical given the amount of time that had passed between students taking the M-CBM in seventh grade and the NC EOG Math Test in eighth grade. As stated in research question 5, this result might suggest that the school did not use these data for instructional planning purposes.

Discussion

The research conducted in this study, which investigated the criterion validity of M-CBM probes at the secondary level, yielded several findings relevant to M-CBM. The first finding indicates that while computation skills significantly correlate with a high-stakes math test in seventh grade, it does not produce a strong correlation. Results indicated that computation skills only accounted for 4% of the variance in NC EOG Math Test scores in both 2012 and 2013 comparisons. This finding indicates that M-COMP has weak concurrent and predictive validity when the NC EOG Math Test is used as a criterion.

Computation skills in isolation, while important, are inadequate for success even at the elementary level (Fuchs et al., 2008). Despite the fact that conceptual math knowledge is needed early in elementary school, math assessments are mainly focused on measuring students' acquisition of computational skills, which likely have a strong relationship with calculation fluency. While computation skills may be crucial skills in early grades, it does not appear to be sufficient in later grades as students are expected to apply these basic skills to more complex math word problems and multi-step analyses in middle school. Secondary math assessments focus less on basic skill development and more on application skills that are not adequately represented by math calculation skills.

Another significant finding from this research was that M-CAP demonstrated strong concurrent and predictive validity in seventh grade, which is an extension of current literature that has shown concurrent and predictive validity for math application CBM in elementary grades. Results indicated that M-CAP scores accounted for 43% of the variance in NC EOG Math Test scores from both 2012 and 2013. These results indicate that M-CBM containing multi-step problems, reasoning, and calculation skills are well equipped to predict performance on a high stakes math test. The strong correlations between the M-CAP and the NC EOG Math Test provide evidence that M-CAP is an appropriate M-CBM for measuring math skills at the secondary level.

Limitations of the Study

Participants in this study included seventh and eighth grade students who took M-CBM during their seventh grade year (2011-2012) and also took the NC EOG Math Test in 2011-2012 and 2012-2013. Students taking a modified version of the NC EOG Math Test, also known as the EXTEND 2, were not included in this study. Their exclusion, while intentional, may have skewed the results of this study as the lowest performing students with learning disabilities or intellectual disabilities were not included in the analysis. Further, both samples contained less than 7% of students with disabilities; therefore, the potential uses of the present study may not be applicable to students with such difficulties. In addition, approximately 71% of students included in this study were identified as White/Non-Hispanic. The remaining 29% of the sample were comprised of students from African American, Hispanic, and multi-racial backgrounds. It is important to note that while some Hispanic students were included in this study, their minimal representation may not be sufficient to suggest that the results of this study are generalizable to students from these ethnic backgrounds. In addition, the lack of representation from students of Asian backgrounds is problematic for overall generalizability. As a result, replications are recommended in settings that have higher representation of racial and ethnic minorities to ensure the ability to generalize the findings to larger populations.

It is also worth noting that because of the archival nature of the study no data were collected regarding the fidelity of the administration procedures. This includes the administration protocol for the NC EOG Math Test and the M-CBM assessments. Therefore, no evidence was obtained about the integrity of the data collected. For this reason information obtained in the study should be interpreted cautiously.

A limitation applies to the interpretation of research questions 5 and 6. Due to North Carolina changing the standards, the 2012 NC EOG Math Test is a different test than the 2013 NC EOG Math Test. However, even though the tests are different, the North Carolina Department of Public Instruction has provided data to suggest that tests are very similar and still significantly correlate with one another. Also, data from the current study show very little difference between concurrent and predictive validity.

Limitations also exist due to the variation in student maturity and quality of instruction over the course of one year. As stated previously, it remains unclear if the data from the M-CBMs influenced the type of instruction students received during their eighth grade year. Similarly, results could also be different in other districts that use different curricula and/or instructional practices. Generalization to other states could also be problematic given that different states have shown significantly different levels of difficulty on their own individual state assessments (Kingsbury et al., 2003).

Recommendations for Future Research

Future research is needed to validate and expand upon the results of this study. Replications are suggested to examine the consistency of these findings across a variety of settings and among different populations as studies with more representative and larger populations may help to obtain a sample that is more representative of all skill levels. It is also important for this research to be conducted across the United States as statewide assessments are designed based on standards that are selected by each individual state. As shown by the findings of Kingsbury et al. (2003), state assessments have significant limitations not limited to their reliability and validity. Significant differences between test content across states hinders the use of a state assessment as a valid and reliable measure of student achievement. Despite this, these assessments continue to be required by federal legislation (NCLB, 2001) and remain an important standard on which students, teachers, and school districts are evaluated. Replication of this research in other states will continue to help validate the psychometric properties of M-CBM and will, therefore, help determine the criterion validity of these probes even when the criterion is different. Also replication of the study using a nationally normed test such as the National Assessment of Educational Progress (NAEP) has the potential to provide further information about criterion validity of M-CBM. Comparison to the NAEP is desirable because it removes the variability that exists between different states and their math assessments. Also comparisons to other standardized, group-administered math tests could provide even further evidence for the criterion validity of M-CBM.

In addition, replication of this study on a larger scale across different areas of the United States will provide a more diverse racial background of students. The inclusion of more Hispanic and Asian students is crucial to improving the validity of this research as these students make up the majority of the population of many school districts in the United States. The same can be stated for students with disabilities. Finally, studies that account for the fidelity and integrity of the administration procedures are needed as adherence to these protocols could make a significant impact on performance.

An expansion of this study should also include a longitudinal approach to research. Administering M-CBM throughout the school year would allow data to be gathered on the reliability of a student's rate of improvement (slope) over the course of a school year. These data would expand the use of the M-CBM from a screening measure to a measure that could be used reliably to monitor student progress at the secondary level.

It may be beneficial to replicate this study using the new AIMSweb M-COMP probes as new scoring procedures have recently been developed. AIMSweb has sought to improve mathematics measurement across grades. In their most recent publication of math probes, AIMSweb has changed the scoring procedure for M-COMP from a digits correct procedure to one that is similar to the scoring procedures for the M-CAP (Pearson, 2012). This scoring procedure is based more on giving points depending on whether the entire problem is correct. The use of this new scoring protocol may result in higher concurrent and predictive validity of M-COMP probes when a state assessment is used as the criterion measure.

Currently, very little research has been conducted studying the connection between math and reading assessments. Specifically, little is known about the relationship between reading CBM and M-CBM (Rutherford-Becker & Vanderwood, 2009). Students who struggle in reading and students who struggle in both math and reading, have the same growth trajectory as it pertains to reading (Jordan, Hanich, & Kaplan, 2003). Thus, math skills do not appear to significantly influence reading progress; however, some aspects of math require reading skills as in solving a math word problem (Jordan et al., 2002). Thurber et al. (2002) demonstrated that reading skills are significantly correlated not only with math applications but also with math calculation skills. This study demonstrated that reading skills correlated with math at almost the same level that separate math skills correlated with each other. Regression models also indicated that when reading ability was part of a prediction model, prediction accuracy increased significantly. Findings suggest that reading could be an essential part of math ability and should not ignored when considering overall math skills (Thurber et al., 2002).

A study completed by Rutherford-Becker and Vanderwood (2009) demonstrated that a reading comprehension CBM could predict applied math performance. This has several implications as it suggests that applied math tests are not only measuring math skills, but also reading skills. As a result, students could be proficient in math, but not perform well on math tasks due to difficulties in reading. This research is important in helping to guide instructional planning for educators as an increase in reading comprehension could also increase performance on applied math tasks.

Overall, more research is needed about the unique contribution of reading skills to math performance, especially at the secondary level. The link between reading and math skills appears undeniable and future studies should incorporate reading skills into prediction analysis.

Implications for the Practice of School Psychology

The applications of CBM are wide as they are used for a variety of purposes including universal screening, progress monitoring, program evaluations, and special education determination. Unfortunately, their validity in each of these applications is also varied. With the further expansion of MTSS models into secondary settings, school psychologists will need access to instruments that demonstrate adequate psychometric properties across their many applications.

School psychologists can play a vital role in helping districts select appropriate tools and interpret data from universal screening measures. Using instruments with good psychometric properties will help to ensure that sound decision practices are used to determine how to best meet student needs. For example, the current study demonstrates that AIMSweb M-CAP probes have very good concurrent and predictive validity, thus adding to the arsenal of instruments that can be used for the purpose of universal screening at the secondary level in the area of mathematics.

Also, practitioners need to understand that calculation skills, while important, do not necessarily equate to overall math knowledge. Math ability at the secondary level appears to be a conglomeration among many distinct abilities including, but not limited to, reasoning skills, calculation skills, and reading ability. Math application M-CBM appears to be the better assessment for students at the secondary level. However, math calculation CBM may be best used as a part of a survey level assessment for students at the secondary level. This can be completed as part of the multi-disciplinary problem-solving process on a group of students who perform poorly on a math application CBM. It may be more feasible and practical for a small group of students to be given a calculation probe to help decipher which students are lacking basic computation skills and which students are struggling with math application skills. A plan can then be designed to help meet the needs of those particular students.

Summary

The results of the current study highlight two important findings. First, M-COMP probes have questionable concurrent and predictive validity at the secondary level. This suggests computation skills are inadequate at the secondary level to make judgments about overall student math ability. Second, M-CAP probes did show both concurrent and predictive validity when the NC EOG Math test is utilized as a criterion measure. This indicates these probes may have use at the secondary level in identifying students who may be in need of further academic support within the area of mathematics.

Upon completion of the analyses described above, the limitations of this study were discussed. These limitations included homogeneous populations, exclusion of the lowest performing students, unknown fidelity and integrity of assessment procedures, maturation of the sample, and multiple treatment interference. Many of these limitations were addressed as recommendations for future research. Additionally, recommendations for the application of this study to the practice of school psychology were presented. These recommendations suggest that school psychologists should understand the psychometric properties of the CBM they use. This awareness and knowledge will help ensure that the best possible recommendations are made when assisting with the interpretation of individual, school, and district data.

References

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance.* Los Angeles, CA: UCLA Center for the Study of Evaluation/National Center on Research Evaluation, Standards, and Student Testing.

AIMSweb. (2008). *AIMSweb training workbooks for early reading, reading, spelling, writing, early numeracy, and math*. Eden Prairie, MN: Edformation. Retrieved from http://AIMSweb.com

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based Effects of acceptability on teachers' implementation of curriculum-measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*, 113-120.

Alonzo, J., & Tindal, G. (2009) *Teachers' manual for regular easyCBM: Getting the most out of the system*. Eugene, OR: University of Oregon. Retrieved from http://www.EasyCBM.com

Alonzo, J., & Tindal, G. (2011) *Teachers' manual for regular easyCBM: Getting the most out of the system*.

Eugene, OR: University of Oregon. Retrieved from http://EasyCBM.com

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Averill, O. (2011). Multi-tier system of supports. *District Administration*, *47*(8), 91.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., . . . Reschly, D. (2005). IDEA 2004 and *response to intervention: policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.

Bazemore, M. (2008). *The North Carolina mathematics tests edition 3: Technical report.* Retrieved from North Carolina Department of Public Instruction website: http://ncdpi.edu

Braden, J., & Tayrose, M. (2008). Best practices in educational accountability: High stakes testing and educational reform. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 575–588). Bethesda, MD: National Association of School Psychologists.

Braden, J. P., & Schroeder, J. L. (2004). High stakes testing and No Child Left Behind: Information and strategies for educators. In A. Canter, S. Carroll, L. Paige, M. D. Roth, & I. Romero (Eds.). *Helping children at home and school* (2nd ed., pp. 73-77). Silver Spring, MD: National Association of School Psychologists.

Breakwell, G. M. (2006). *Research methods in psychology* (3rd ed.). Thousand Oaks, CA: Sage.

Brualdi, A. (1999). Traditional and modern concepts of validity. *ERIC Clearinghouse on Assessment and Evaluation*. Retrieved from www.ericdigests.org/2000-3/validity.htm

Busch, T. W., & Espin, C. A. (2003). Using curriculum-based measurement to prevent failure and assess learning in the content areas. *Assessment for Effective Intervention, 28*, 49–58. doi:10.1177/073724770302800306

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and*

Special Education, 24, 235-245. doi:10.1177/07419325030240040601

Cawley, J., Parmar, R., Foley, T. E., Salmon, S., & Roy, S. (2001). Arithmetic performance of students: Implications for standards and programming. *Exceptional Children*, 67, 311–328.

Chubb, J. E. (Ed.). (2005). *Within our reach: How America can educate every child*. New York, NY: Rowman & Littlefield.

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

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

Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olsen, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review 36*, 562-581.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184–192. doi:10.1177/00224669030370030801

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507–524.

Dworkin, A. G. (2005). The No Child Left Behind Act: Accountability, high-stakes testing, and roles for sociologists. *Sociology of Education, 78*, 170-174. doi:10.1177/003804070507800205

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & 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*, 247–265.

Espin, C., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in the content areas: Vocabulary matching as an indicator of progress in social studies learning. *Journal of Learning Disabilities, 38*, 353–363. doi:10.1177/00222194050380041301

Evans, J. D. (1996). *Straightforward statistics for the* *behavioral sciences*. Pacific Grove, CA: Brooks/Cole.

Evans-Hampton, T. *N*., Skinner, C. H., Henington, C., Sims, S., & McDaniel, E. (2002). An investigation of situational bias: Conspicuous and covert timing during curriculum-based measurement of mathematics across African American and Caucasian students. *School Psychology Review, 31*, 529-539.

Foegen, A. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *Journal of Special Education*, *35*, 4-16. Retrieved from Education Research Complete database.

Foegen, A. (2008). Progress monitoring in middle school mathematics: Options and issues. *Remedial and Special Education*, *29*, 195-207. doi:10.1177/0741932507309716.

Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique, 25*, 175-203.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *Journal of Special Education, 35*, 4-16.

Foegen, A., Olson, J., & Impecoven-Lind, L. (2008). Developing Progress Monitoring Measures for Secondary Mathematics. *Assessment for Effective*

Intervention, *33*, 240-249. doi:10.1177/1534508407313489

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measuring in mathematics: A review of the literature. *The Journal of Special Education*, *41*, 121-139. doi:10.1177/00224669070410020101.

Fuchs, L. S., Fuchs, D., Zumeta, R. O. (2008). A curricular-sampling approach to progress monitoring: Mathematics Concept and Applications. *Assessment for Effective Intervention, 33*, 225-233. doi:10.1177/1534508407313484

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Dutka, S., & Katzaroff, M. (2000). The importance of providing information on the structure and scoring of performance assessments. *Applied Measurement in Education, 13*, 1-34.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, *R*., & Jancek, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology*, *95*, 293-305. doi:10.1037/0022-0663.95.2.293

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*, 263–279.

Fuchs, D., Mack, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18*, 157–171.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between interventionally relevant measurement models. *Exceptional Children, 57,* 488–501.

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. *N*. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research & Practice, 22*, 13-24. doi:10.1111/j.1540-5826.2007.00227

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kupek, P.,Stecker, P. M. (1994). Technical features of a mathematics concepts/applications curriculum-based measurement system. *Diagnostique, 19*, 23-49. doi:10.1177/002246690704100201 01

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1999). *Monitoring basic skills progress: Basic math manual*. Austin, TX: PRO-ED.

Gaur, A. S., & Gaur, S. S. (2006). *Statistical methods for practice and research: A guide to data analysis using SPSS*. Thousand Oaks, CA: Sage.

Grimm, K. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neural Psychology, 33*, 410-426. doi:10.1080/87565640801982486

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education, 36*, 112–112. doi:10.1177/00224669020360020501

Helwig, R., Heath, B., & Tindal, G. (2000). *Predicting middle school mathematics achievement using practical and efficient measurement instruments*. Retrieved from University of Oregon, Behavioral Research and Teaching website: http://brt.uoregon.edu/pub/lsassess.html

Hintze, J., Christ, T., & Methe, S. (2006). Curriculum-based assessment. *Psychology in the Schools*, *43*, 45-56. doi:10.1002/pits.20128.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*, 514-528.

Hoover, H., Dunbar, S., & Frisbie, D. (2005). *Iowa tests of basic skills.* Chicago, IL: Riverside Publishing Company.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement.* New York, NY: Guilford.

Hudson, *P.*, & Miller, S. P. (2006). *Designing and implementing mathematics instruction for students with diverse learning needs.* Boston, MA: Allen & Bacon.

Ikeda, J. M., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 103-114). Bethesda, MD: National Association of School Psychologists.

Individuals with Disabilities Education Improvement Act of 2004, Public Law No. 108-446 (2004). Retrieved from www.gpo.gov/fdsys/pkg/BILLS-108hr1350enr.pdf

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

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate production scores. *School Psychology Review, 34*, 27–44.

Jitendra, A. K., Sczesniak, E., & Deatline-Buchman, A. (2005). An Exploratory Validation of Curriculum-Based Mathematical Word Problem-Solving Tasks as Indicators of Mathematics Proficiency for Third Graders. *School Psychology Review*, *34*, 358-371.

Johnson, B., & Christensen, L. (2004) *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed.). Boston, MA: Pearson Education.

Jordan, *N*. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850. doi:10.1111/1467-8624.00571

Jordan, N. C., Kaplan, D., Hanich, L. B. (2002). Achievement growth in children with learning difficulties in Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94*, 586-597. doi:10.1037/00220663.94.3.586

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

Kelley, B. (2008). Best practices in using curriculum-based evaluation and math. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 419-437). Bethesda, MD: National Association of School Psychologists.

Kingsbury, G. G., Olson, A., Cronin, J., Hauser, C., & Houser, R. (2003). *The state of state standards: Research investigating proficiency levels in fourteen states*. Lake Oswego, OR: Northwest Evaluation Association.

Kovaleski, J., VanDerHeyden, A., & Shapiro, E. (2013). *The RTI approach to evaluating learning disabilities*. New York, NY: Guilford Press.

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy* (NCES 2006-483). Washington, DC: U.S. Department of Education, National Center for Education Statistics

Levine M. D., & Reed M. (1999). *Developmental variation and learning disorders.* Educators Publishing Service: Cambridge and Toronto.

McCook, J. E. (2006). *The RTI guide: Developing and implementing a model in your schools.* Horsham, PA: LRP Publications.

Merrell, K. W., Ervin, R. A., & Gimpel, G. A. (2006). *School psychology for the 21st century: Foundations and practices.* New York, NY: Guilford Press.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher 18,* 5-11.

Messick, S. (1990). Validity of test interpretation and use (ETS-RR-90-11). Educational Testing Service. Princeton, New Jersey.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards. Washington, DC

National Center for Education Statistics. (2014). *School district demographics system*. Retrieved from http://nces.ed.gov/globallocator/

National Center for Education Statistics. (2009). *The nation's report card: Mathematics 2009* (NCES 2010-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA:

Ott, L., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis* (6th ed.). Belmont, CA: Cengage Learning.

Pearson (2012). *AIMSweb: Technical Manual.* Bloomington, MN: Pearson.

Preacher, K. J. (2002). Calculation for the test of difference between two independent correlation coefficients [Computer software]. Available from http://quantpsy.org

The No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2001). Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Reschly, D. J. (2008). Paradigm shift and beyond. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 3–15). Washington, DC: National Association of School Psychologists.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences, 17,* 147–159. doi: 10.1016/j.lindif.2007.03.010

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. *California School Psychologist, 14*, 23–34. doi:10.1007/BF03340948

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2009). *Assessment: In special and inclusive education.* Belmont, CA: Wadsworth, Cengage Learning.

Scafidi, T., & Bui, K. (2010). Gender similarities in math performance from middle school through high school. *Journal of Instructional Psychology*, *37*, 252–255.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical Adequacy of the maze task for curriculum-based measurement of Reading Growth. *The Journal of Special Education, 34,* 164-172. doi:10.1177/002246690003400305

Shinn, M. R. (2008). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 243–261). Washington, DC: National Association of School Psychologists.

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

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

Sibley, D., Biwer, D., & Hesch, A. (2001). *Establishing Curriculum-Based measurement oral reading fluency performance standards to predict success on local and state tests of reading.* Retrieved from ERIC database. (ED453527)

Stecker, P., Fuchs, L., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*, 795–819. doi:11.1102/pits.20113

Sugai, G., Horner, R. H., & Gresham, F. M. (2002). Behaviorally effective school environments In M. R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 315–350). Bethesda, MD: National Association of School Psychologists.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology, 89*, 276-288. doi:10.1037/00220663.89.2.276

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*, 498-513.

Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17-36). Bethesda, MD: National Association of School Psychologists.

Tilly, W. D., Reschly, D. J., & Grimes, J. (1999). Disability determination in problem-solving systems: Conceptual foundations and critical components. In *Special education in transition: Functional and noncategorical programming.* Longmount, CO: Sopris West.

Tindal, G., Germann, G., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Report No. 132). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Twyman, T., & Tindal, G. (2007). Extending curriculum-based measurement into middle/secondary schools: The technical adequacy of the concept maze. *Journal of Applied School Psychology, 24*, 49–67. doi:10.1300/J370v24n01_03

Wallace, T., Espin, C. A., McMaster, K., Deno, S. L., & Foegen, A. (2007). CBM progress monitoring within a standards-based system. *The Journal of Special Education, 41*, 66-67. doi:10.1177/00224669070410020201

Appendix A - Permission from Internal Review Board

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

April 5, 2013

Giancarlo Anselmo
814 Brooklee Drive
Kings Mountain, NC 28086

Dear Mr. Anselmo:

Your proposed research project, "Predictive Validity of M-CBM at the Secondary Level," (Log No. 13-065) has been reviewed by the IRB and is approved as an expedited review for the period of April 4, 2013 to April 4, 2014.

It is also important for you to note that IUP adheres strictly to Federal Policy that requires you to notify the IRB promptly regarding:

1. any additions or changes in procedures you might wish for your study (additions or changes must be approved by the IRB before they are implemented),
2. any events that affect the safety or well-being of subjects, and
3. any modifications of your study or other responses that are necessitated by any events reported in (2).

Should you need to continue your research beyond April 4, 2014 you will need to file additional information for continuing review. Please contact the IRB office at (724) 357-7730 or come to Room 113, Stright Hall for further information.

Although your human subjects review process is complete, the School of Graduate Studies and Research requires submission and approval of a Research Topic Approval Form (RTAF) before you can begin your research. If you have not yet submitted your RTAF, the form can be found at http://www.iup.edu/page.aspx?id=91683 .

This letter indicates the IRB's approval of your protocol. IRB approval does not supersede or obviate compliance with any other University policies, including, but not limited to, policies regarding program enrollment, topic approval, and conduct of university-affiliated activities.

I wish you success as you pursue this important endeavor.

Sincerely,

John A. Mills, Ph.D., ABPP
Chairperson, Institutional Review Board for the Protection of Human Subjects
Professor of Psychology

JAM:jeb

Cc: Dr. Joseph Kovaleski, Dissertation Advisor
Ms. Brenda Boal, Secretary

Appendix B - Extension from Internal Review Board

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the
Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

May 28, 2014

Giancarlo Anselmo
814 Brooklee Drive
Kings Mountain, NC 28086

Dear Mr. Anselmo:

Your request for continuing review for your research project, "Predictive Validity of M-CBM at the Secondary Level," (Log No. 13-065), has been reviewed by the IRB and is approved as an expedited review for the period of May 27, 2014 to May 27, 2015.

<u>You should read all of this letter, as it contains important information about conducting your study</u>.

Now that your project has been approved by the IRB, there are elements of the Federal Regulations to which you must attend. IUP adheres to these regulations strictly:

1. You must conduct your study <u>exactly as it was approved</u> by the IRB.
2. <u>Any additions or changes</u> in procedures <u>must</u> be approved by the IRB <u>before</u> they are implemented.
3. You must notify the IRB promptly of <u>any</u> events that affect the safety or well-being of subjects.
4. You must notify the IRB promptly of any modifications of your study or other responses that are necessitated by any events reported in items 2 or 3.

Should you need to continue your research beyond May 27, 2015 you will need to file additional information for continuing review. Please contact the IRB office at (724) 357-7730 or come to Room 113, Stright Hall for further information.

The IRB may review or audit your project at random *or* for cause. In accordance with IUP Policy and Federal Regulation (45CFR46.113), the Board may suspend or terminate your project if your project has not been conducted as approved or if other difficulties are detected.

It is strongly recommended that all researchers and their advisors complete CITI on-line protection of human subjects and responsible conduct of research training. The training is available at http://www.iup.edu/page.aspx?id=93408 and there is no charge to you.

I wish you success as you pursue this important endeavor.

Sincerely,

Jen Roberts

Jennifer J. Roberts, Ph.D.
Chairperson, Institutional Review Board for the Protection of Human Subjects
Professor of Criminology

JJR:jeb

Cc: Dr. Joseph Kovaleski, Dissertation Advisor