UNIVERSITY OF WISCONSIN-LA CROSSE

Graduate Studies

CURRICULUM-BASED MEASUREMENT IN MATHEMATICS: PREDICTING FUTURE PERFORMACE ON STATE ASSESSMENTS

A Chapter Style Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Education Specialist

Katherine Ann Stinson, M.S.E.

College of Liberal Studies
Psychology/School Psychology

May, 2011

CURRICULUM-BASED MEASUREMENT IN MATHEMATICS: PREDICTING

FUTURE PERFORMACE ON STATE ASSESSMENTS

By Katherine A. Stinson

We recommend acceptance of this thesis in partial fulfillment of the candidate's requirements for the degree of Education Specialist.

The candidate has completed the oral defense of the thesis.

2-4-11
Robert J. Dixon, Ph.D. Thesis Committee Chairperson	Date
	2-4-11
Carol A. Angell, Ph.D. Thesis Committee Member	Date
	2-4-11
David M. Reineke, Ph.D. Thesis Committee Member	Date

Thesis accepted

Vijendra K. Agarwal	3/16/11
Vijendra K. Agarwal, Ph.D. Associate Vice Chancellor for Academic Affairs	Date

ABSTRACT

Stinson, K.A. Curriculum-based measurement in mathematics: Predicting future performance on state assessments. MS in Education, December 2010, 74pp. (R. Dixon).

Curriculum-Based Measures (CBMs) in reading and their relationship to state standardized tests have been established and an important feature of CBMs. This study will examine the predictive relationship of CBMs (i.e. computation and application) in mathematics on a state standardized assessment for elementary students. This will assist educators in the early identification of problems and the adjusting of their instructional practices in order to promote proficiency on state assessments. The current study used 97 students from one school (35 third graders, 25 fourth graders, and 37 fifth graders). Students were given three probes, two weeks apart. The average scores on the probes tended to increase in all three grade levels for both computation and concepts/applications.   A regression analysis indicated that both computation and application probes were significant in predicting WKCE results, indicating that math CBMs can be a valid tool in predicting future performance on WKCEs.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

While there is the age-old concern for teaching children to be good readers, there is a growing concern about the math literacy of American children. Currently, only 30-40% of all students are performing at the "proficient" level or above in mathematics as measured by the National Assessment of Educational Progress (NAEP; Lee, Grigg, & Dion, 2007). Additionally, with the implementation of No Child Left Behind (NCLB) 2001, there is emphasis placed on statewide achievement testing as a measure of accountability for school achievement. The passage of NCLB (2001) requires that all states use some form of statewide achievement test to determine accountability for student progress. The standardized tests being used by school districts to demonstrate student performance outcomes are most often based on curriculum standards established by the state (Braden & Tayrose 2008; Erikson, Ysseldyke, Thurlow & Elliot, 1998; Linn, 2000). This low performance level on the NAEP reflects gaps between students' performances on the Adequate Yearly Progress (AYP) that school districts are expected to make.

With the passage of NCLB standards have been selected by states that hold school districts accountable for the learning of their students. Schools must provide proof that students are learning and demonstrating academic growth through the school year. While accountability is a positive attribute of NCLB, the implications of this act and the accountability expectations that are associated with it result in the possibility of financial sanctions for schools that are unable to have students attain AYP. Rather than allow the children to wait and fail on these standardized assessments, school officials are moving to establish assessment systems that allow for early identification and the accurate prediction of future student performance to ensure that all students are meeting stakeholder expectations by increasing the levels of academic proficiency (Braden & Tayrose, 2008). While there are currently assessment systems that allow for this type of identification in reading, such as the use or oral reading fluency (ORF) Curriculum Based Measurements (CBMs) to predict reading performance, educators and administrators are seeking an efficient tool to gauge current student classroom math performance and predict future performance on state testing in mathematics. Many researchers suggest that statewide standardized testing could be supplemented with a progress monitoring system measuring student achievement and document growth towards the state standard (Deno, 2003; Jiban & Deno, 2007). They suggest that by using CBMs, it may possible to use the data in order to make immediate instructional changes, therefore providing the necessary changes to positively adjust the curriculum to meet the needs of the students before a failing score is recorded (Crawford, Tindal, & Steiber, 2001; Deno, 2003; Jiban & Deno, 2007; Shapiro et al., 2006; Stecker, Fuchs, & Fuchs, 2005).

A further complication to schools educating students to meet levels of proficiency is the schedule of statewide testing and the feedback to the teachers. Students take these statewide exams once a year. Often, the results are not reported until near the end of the school year, giving both teachers and students minimum feedback through the duration of the school year as to level of proficiency and the instructional needs and impact of instructional activities. In addition, the growth in students from the assessment to the reporting of the results means that the assessment results are not meaningful to the provision of services for next year. Therefore, there is a lack of data and time to adjust the curriculum for those students who are struggling with classroom content. Educators may be uncertain about the student's performance on statewide-standardized tests until the end of the school year, and at that time, it is often too late to teach the lost information. Therefore educators would greatly benefit from a tool that would help predict performance on statewide testing, allowing them to modify curriculum earlier in the school year for the students who are struggling to master the content and thus have difficultly to score in the proficient range. CBMs could be an objective means of early identification of those students who are experiencing difficulties in the classroom as well as determine who is benefiting from the instruction. They offer schools a way to attend to student progress and instructional methods and programs. CBM is a tool that tracks a student's rate of progress toward a particular goal (Deno, 1985, 2003; Shinn, 1989).

The goal of this research study is to determine the viability of math CBMs to predict performance on a future statewide assessment. The original intent of CBMs was for teachers to have a technically sound, simple method to document student growth and determine the need to modify instructional programs. The hope was that by modifying instruction in response to patterns of poor performance a teacher could improve student achievement (Deno, 1985). Recently, research has demonstrated the use of CBM data to predict success on high stakes assessment and to measure growth in content areas of reading fluency, mathematics, and writing (Deno, 2003). If CBMs can be extended to the statewide tests and accountability expectations, the impact of this study will be on the opportunity for teachers to extend their interventions to make an impact in the classroom as well as positively impact the accountability that is expected in schools today.

The review of literature review will begin with a discussion on high stakes testing, how CBMs fit into the school system and the RtI model, an overview of CBMs including the features and different types, a discussion of accountability, CBMs in predicting reading performance, our current math expectations, and using math CBMs to predict state standardized testing

High Stakes Testing

Many states and school districts are required to administer tests in order to gather data about student achievement, and to hold schools accountable. Tests are considered to be "high stakes" if the results have serious consequences for students or educators (American Educational Research Association, 2000). Many states have tests that are given yearly to students in elementary, middle, and high schools. These tests serve several purposes. They measure how each student has achieved in comparison to a state or national standard, or level of knowledge and that all students are expected to have learned to their a grade level proficiency. These test scores of students' academic knowledge are then used to determine the effectiveness of instruction. Therefore, low scores have negative implications for many schools. While the sanctions depend on funding and the individual state, a few key principles are constant: low scoring schools are more likely to lose autonomy over funds (Braden & Tayrose, 2008). According to NCLB law, sanctions are federally mandated for schools receiving NCLB funding to hold schools accountable for student progress (No Child Left Behind Act of 2001, as cited in Braden & Tayrose). If a school receiving NCLB funds fails to meet AYP goals two years in a row, students in the district have the option of school choice in that they may choose to attend a school in another district at the school's expense. If AYP is not met after three years, schools must provide supplemental educational services such as tutoring. After year four, the school must adopt scientifically based methods of instruction and train all teachers in these methods. After five years, the school and government must develop a plan for external contracting, and after six years, the school gets closed and may reopen as a charter school, or other alternative school (Braden & Tayrose, 2008). Students, schools, personnel, and the state have a great deal at stake with this testing. For this reason, the term "high-stakes testing" is often used to describe these tests (McLane, 2001).

This type of accountability within the educational system places a great deal of emphasis on teachers who are responsible for assisting students in reaching proficient levels on standardized testing. Pollock (2007, p.18) states that, "teachers are the primary factor that affects individual student success." When teachers use "the big four": a well articulated curriculum, plan for delivery, vary assessment, and give criterion-based feedback, success in student achievement is often heightened (Pollock, 2007).

Curriculum Based Measurements and High Stakes Testing

A variety of studies have found high correlations between CBM scores on reading and achievement levels on high-stakes tests. In addition, studies have found a relationship between CBM reading scores and pass rates on the statewide tests in a number of different states. Research in Oregon has showed that students who, at the beginning the third grade, can read more than 110 words correctly in one minute are likely to pass the Oregon statewide test given in third grade (Good, Simmons, & Kameenui, 2001;

McLane, 2001). Eighth grade students who can read 145 words per minute from local newspaper passages in 1 minute are almost certain to pass the Minnesota Basic Skills Test in reading (Muyskens & Marston, 2002). Powell-Smith (2004) reported on results of all published studies that examined relationships between CBM measures of reading (oral reading fluency [ORF]) and outcomes of statewide assessments. He found data in 8 states with average correlations between ORF and performance on state testing to be .60-.75 range, indicating a strong relationship between how a student performs on a reading CBM and their performance on state standardized testing. There is little information regarding math CBMs and performance on statewide tests.

There have been only three published articles on performance on math CBMs and outcomes on state standardized tests in Oregon, Pennsylvania, and Minnesota (Helwig, Anderson, & Tindal, 2002; Jiban & Deno, 2007; Shapiro, et al., 2006). These studies found the correlations between math CBMs and state-testing performance to be somewhat lower than reading and range between .50-.65. Jiban & Deno's study included a sample size of 35 third grade students and 49 fifth grade students. The correlation between math CBMs and state testing was a range between .11-.44. They discussed that the sample size was inadequate and had an effect on their reliability and validity of the study. Additionally, they encouraged further research to extend the sample size to include all students to produce more scientifically sound results. Shapiro and colleagues report that their results, a correlation range of .50-.60 between math CBMs and state testing, are limited to Pennsylvania and additional information is needed from other states to confirm that math CBMs are a powerful predictor of statewide testing. Finally, Helwig, Anderson, & Tindal limited their results, an 87% prediction rate of meeting state standards when comparing CBMs to state testing, to examining math performance in middle school students with learning disabilities. Therefore, additional research is needed in the elementary school level as well.

The current study will focus on how students in grades 3-5 perform on both computation and application math CBMs and how their performance will correlate with state standardized testing in mathematics.

The New Role of a School Psychologist

Traditionally, a school psychologist has been referred as the "gatekeeper" of special education (Bradley-Johnson & Dean, 2000). A school psychologist's primary role was to test students and use the testing information to decide whether or not the student meets criteria for special education. The role of a school psychologist is experiencing a shift in the nature of the identification of students with learning disabilities. As early as 20 years ago, educators have noted that there is bias in the role of referral and identification in special education (Marston, Mirkin, & Deno, 1984). The current discrepancy model, often referred to as "wait to fail or refer-test-place" requires a student to be performing poorly in the classroom for an extended of period of time in order for a psychoeducational assessment to take place (Merrell, Ervin, & Gimpel, 2006). The next step is a referral made by the teacher and the parent consenting to an evaluation. A school psychologist then administers an intelligence (IQ) test and an achievement test. This allows school psychologists to determine if a discrepancy exists between the student's intelligence and achievement. Only then, is a student eligible to be provided with the special education services. Those students who perform poorly in school, but do not meet the requirement of the ability-achievement discrepancy are usually not eligible to receive services for a specific learning disability.

The new framework for the role of school psychologists is Response to Intervention (RtI). This model requires a three-tier system to intervention. Tier I consists of core instruction in which 80-85% of students are meeting expectations, or benchmarks (Reschly, 2008). This is where CBMs can first be implemented, and demonstrates why they are an important tool in assessing student performance. In tier I, CBMs are used for universal screening. The measures are generic, simple, and time efficient, which makes them an excellent tool to use for every student. By administering CBMs and using the norms of the classroom as a comparison point, teachers can identify the students who are “at risk” for poor classroom achievement (Deno, 1986). If a student is not successful on this universal screening, it may suggest that they are not benefitting from the core curriculum and may need to advance to tier II. Tier II typically consists of small group or individual interventions such as tutoring (Reschly, 2008). CBMs are still used at this level, primarily in progress monitoring to determine if the student is able to make academic progress with the additional instructional interventions at the tier II level or if it is necessary to move to tier III. Tier III involves more intensive and individualized instruction than tier II. Sometimes, tier III placement can lead to special education placement, but not necessarily (Reschly). In this stage CBMs or other progress monitoring tools are used more frequently, perhaps weekly, in order to determine the effectiveness of the general and supplemental instruction. By determining the effectiveness of the instruction at the various levels and intensity of instruction, educators can use the information provided by the CBMs and modify the instruction based on the need of the student (Reschly).

The main concern with the current practices of the discrepancy model is that the process seems to require one to wait for a problem to manifest to a point where the student is failing. The new focus is to prevent a student from reaching the point of failure, for many reasons. The services a school psychologist provides for a child are both time consuming and expensive with potentially little to show for it (Shinn, 1989). Psychoeducational assessments can cost thousands of dollars and take days to complete. Chambers and colleagues (2004) determined the recent costs of per student expenditures in special education ranges from $10,558 for students with learning disabilities to a high of $20,095 for students with more severe or multiple disabilities. They found the average cost per student for a regular education is $6,556. This data indicates the average cost for students with learning disabilities is 1.6 times that of a general education student (Chambers, Parish, & Harr, 2004)

Another downfall of the traditional discrepancy model is unnecessary referrals that result in a lot of formal testing of a student (Marston, Mirkin & Deno, 1984) Sometimes referrals are made for students based on behavioral concerns that manifest in the academic area, but are not necessarily demonstrating a significant skill discrepancy. CBMs allow for clear academic data demonstrating a students' progress or lack thereof. The general guideline established in reading by Marston, Mirkin, and Deno is that students who are performing 50% lower than the average of the class should receive additional academic interventions. Therefore, CBMs save time in unnecessary assessments and also provide baseline data that can be used for IEP goal setting if special education services are needed (Deno, 1986).

History of Curriculum Based Measurements

CBMs are a tool that is becoming increasingly popular in the educational field as a tool of assessment (Shinn, 2008). CBMs originated in the mid to late 1970s at the University of Minnesota's Institute for Research on Learning Disabilities (IRLD). Stan Deno and his colleagues attempted to create a tool to assist special educators in tracking educational growth, mainly in the area of reading. There were two primary goals in the development of CBMs. First, researches intended to develop a tool that special education resource teachers could use for evaluation of student progress towards individualized education program (IEP) goals (Deno, Mirkin, & Chiang, 1982). The second goal of the research program was to develop a type of measurement and evaluation that teachers could use to make decisions when to modify a students' instructional program (Deno, 1985). This measurement and evaluation program was developed and tested over a period of six years, from 1977 to 1983. The result of the implementation of this tool was that teachers were more effective in their teaching, which was demonstrated by improved student achievement (Deno, 1985; Fuchs, Deno, & Mirkin, 1984).

In the past 25 years, a great deal of research has been done involving reading CBMs including: establishing norms for screening and identifying students in need of special education (Shinn, 1989), evaluating effectiveness of educational programs (Tindal, 1992), reintegrating students with disabilities into general education classrooms (Fuchs, Fernstrom, Reeder, Bowers, & Gilman, 1992), monitoring progress and planning instruction in general education classrooms (Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994), and identifying potential candidates for special education using a dual discrepancy model (Fuchs, Fuchs, & Speence, 2002). While there was a small portion of this research dedicated to writing and math CBMs, the main focus remained on reading.

Types of Curriculum-Based Measures

CBMs are a broad category of measurement used to inform teachers of the effectiveness of their own instruction, and when to modify the instruction if necessary. It displays the skill level of the students and is an indicator of when a student is doing well or struggling (Kelley, Hosp, & Howell, 2008). There are three types of CBMs: General Outcome Measures (GOM), skills based measures (SBMs), and mastery measures (MMs). All are considered efficient in the instructional decision making process (Kelley, Hosp, & Howell, 2008).

GOMs are standardized procedures that sample students' performance using a task that incorporates the use of several sub skills (Kelley, Hosp, & Howell, 2008). GOMs are designed to help teachers and school psychologists examine student performance and progress. They are often used as screening tools, and benchmarks for long-term goals. (Espin & Foegen, 1996; Fuchs & Deno, 1991). Oral reading fluency (ORF) and correct word sequences (CWS) are two kinds of GOM used for reading and writing. Math is frequently not a GOM because math is a multi-skill area where the skills may be related but still independent. Therefore, SBMs are typically used to measure mathematics performance.

The second type of CBM is SBMs, which incorporate a broad variety of skills in one specific domain, such as math computation or math problem solving. SBMs usually sample a cluster of related skills. For example, a computation SBM may include addition, subtraction, multiplication, and division all in one sheet (Kelley, Hosp, & Howell, 2008).

Finally, MMs are direct measures of objectives that are sensitive to changes in learning. They target a very specific skill or skill set and collects a students' performance on this skill. MMs are useful in validating a particular problem area, determining specific skill acquisition, and determining when a student reaches a level of proficiency. An example of MMs is measuring single digit addition, a very specific skill, as opposed to the first two CBMs, which measure more broad skills. (Kelley, Hosp, & Howell, 2008).

Differences between Math and Reading CBMs

When discussing CBMs it is crucial to understand that reading and math are different subjects, therefore, what is successful in regards to reading may not necessarily be applicable to math CBMs. When discussing the different types of CBMs, GOMs are often considered excellent tools in screening and defining targets for long-term goals. They incorporate using and having the student demonstrate several sub skills simultaneously. For example, reading may be broken down into phonemic awareness, fluency, and reading comprehension. The act of oral reading combines all of these skills and it is the proficiency in the global task that allows a teacher to conclude that the sub skills are probably mastered. If, on the other hand, a student does not read well, further testing may be warranted to pinpoint the exact area of concern. Once remediated, the global performance of oral reading will reflect the mastered sub skill along with the other skills that go into reading. However, because math does not involve the use of several sub skills, or a capstone task, developing a GOM for math CBM has been a challenge if not impossible. To date, no single measure communicates generalized or global math skills because math is a multi-skill area (i.e., computation, problem solving). Therefore SBMs have been developed for math to reflect this limitation with the expectation that they can fulfill the function of GOMs and provide a quick and efficient measure that can identify difficulties and relate to important educational outcomes (Kelley, Hosp, & Howell, 2008).

Uses of Curriculum-Based Measures

CBMs typically make use of criterion-referenced measures as opposed to norm-referenced measures. Criterion-referenced measures are used to determine if students are able to reach certain performance levels. It is based on the idea that if students do not know a skill and need additional instruction, they will most likely do poorly on the test of that particular skill (Hosp, Hosp, & Howell, 2007).

CBMs assess student progress towards long-term goals. This means that CBMs are designed to evaluate general outcomes as opposed to mastery of successive objectives. Therefore, in the area of reading, CBMs are often referred to as general outcome measures (Fuchs & Deno, 1991). This is not the case in the area of mathematics. CBMs allow for frequent monitoring and graphing of student scores. They are used in a predictive manner to determine if students are on track to meeting long term goals. Data is also used to judge relative current performance and determine if the instructional practice has been effective in facilitating student growth. This information can be particularly useful in helping teachers individualize instruction. (Stecker, Fuchs, & Fuchs, 2005). The increased sensitivity to showing growth is another key advantage to using CBMs in the classroom. Standardized statewide achievement testing is not sensitive to demonstrating growth as it is only administered once a year (Deno, 1985), and as mentioned previously, relying on this testing does not allow for frequent instructional modifications to fit the needs of the students.

CBMs emphasize repeated measurement over time. Therefore, teachers can generate a rate of progress, or growth. Teachers can use CBM data for progress monitoring to demonstrate the rate of learning as it is occurring. Thus CBMs are a measure of formative assessment, reflecting how a student performs over duration of time. This allows teachers to make immediate adjustments in instruction if the student is not reaching expected growth targets (Hosp, Hosp, & Howell, 2007; Stecker, Fuchs, & Fuchs, 2005). The ability to improve communication of student progress within and between grade levels is a key advantage to the use of CBMs. By measuring current growth and improvement of students over the school year, this information can be passed along as the student progresses through the school system (Deno, 1985).

Features of CBMs

As noted previously, CBMs are composed of several distinct features: technical adequacy, simplicity, and efficiency. Measures can be tied to the curriculum or based on national expectations; students are tested on what they are taught. The content of the CBMs are the same and the stimulus materials should look the same (Hosp, Hosp, & Howell, 2007).

Technical Adequacy

Research has endorsed the use of CBM procedures for measuring ongoing student progress and making instructional decisions (Shinn, 1989). There have been hundreds of studies supporting the application of CBMs in the area of reading (Hosp, Hosp, & Howell, 2007). Reliability has been measured through test-retest, parallel forms, and inter-rater agreement. Test-retest reliability over the span of 1-10 weeks ranged from .82-.97 with most estimates above .90 (Shinn, 1989). In addition, parallel forms were used and correlations ranged from .84-.96, with most correlations above .90 (Shinn, 1989). Inter-rater agreement was also measured and coefficients were .99 suggesting that these measures are consistently scored (Shinn, 1989). Validity was been measured primarily using criterion comparisons. In some of the seminal work regarding CBMs and reading, the relationships between reading out loud and standardized testing ranged from .70-.95 (Deno, 1985). The correlation between words read correctly out loud and number of correct answers on the Literal and Inferential subtests of the Stanford Achievement Test were .78 and .80 respectively (Deno, 1985). The correlation between words read correctly out loud and word identification subtest of the Woodcock Reading Mastery test was .93 (Deno, 1985). As research has continued in this area, the correlations between reading and standardized measures remain in the same high correlation range (Deno, 2003). Much of the seminal research in the area of CBMs focused on reading. While there was some early research conducted in the area of math reliability and validity (Tindal, Germann & Deno, 1983; Tindal, Marston, & Deno, 1983), these studies were not as extensive and comprehensive as studies conducted in the area of reading and also resulted in lower correlations.

Simplicity and Efficiency

In order for teachers to use CBMs regularly in their classrooms, they must be simple to implement in the classroom, as well as an efficient measurement of growth and achievement (Deno, 1985). CBMs are short in duration to allow for frequent administration, available in multiple forms, inexpensive to produce, and sensitive to the improvement of student achievement over time (Shinn, 1989). Additionally, CBMs are often used to share data with others; therefore, standardized procedures must be used. That is, the tasks that are used are consistent across the content area. Standard procedures are followed in selecting or constructing test materials and standard administration and scoring directions exist for each prompt (Hosp, Hosp, & Howell, 2007).

Composition of CBMs

CBMs are typically a compilation of a set of standardized instructions, a timing device, a set of materials (i.e., a reading passage, a set of math problems, etc.), scoring rules, standards for judging performance, and record forms or charts (Hosp, Hosp, & Howell, 2007). The instructions are designed to be straightforward and easy for students to understand. CBMs are designed to be not much different from an activity that a student would engage in during a typical classroom setting. The students are timed in their performance to increase efficiency and so that a score can be obtained and comparisons can be made between students and a current or future standard (Hosp, Hosp, & Howell).

Reading

For CBMs in reading, the most frequently used measure is oral reading fluency (ORF). Although several measures were studied in the Minnesota IRLD, the number of words read correctly in a 1-minute time period produced a reliable and valid indicator of a students' overall reading proficiency (Deno, Mirkin, & Chaiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Marston, 1989). Typically, passages are selected that represent instructional grade level material.

Math

Math CBMs are designed to be easy and efficient to score. They can be administered individually or to an entire class at the same time. Typically, math CBMs are composed of three different areas: Early Numeracy (K-1st grade), Computation (grades 1+), and Concepts and Applications (Grades 2+) (Hosp, Hosp, & Howell, 2007). Depending on the type of probe, they are typically administered from 2-8 minutes and correspond with each grade level (Fuchs, Fuchs, & Courey, 2005). Typically, when scoring, the number of digits correct is what is reported. This is often done because scoring by digits correct as opposed to problems correct is a more sensitive measure of change (Hosp, Hosp, & Howell, 2007). Additionally, it is less biased to score by digits correct because a student is given more credit for solving more complex problems such as multiple digit multiplication (Hosp, Hosp, & Howell, 2007).

Of the few studies conducted in regards to reliability, Math CBMs have demonstrated high inter-rater reliability (.97), high 1-week test-retest (.87), and moderate alternate form reliability (.66) (Tindal, Marston, and Deno, 1983). More recent research in the area of internal consistency found coefficients to be above .80 (Christ et al., 2008). An examination of alternative forms of reliability found coefficients of above .80 as well (Christ et al., 2008).

In comparison to reading CBMs, not much is known about the validity of math CBMs. Content validity is the extent to which a test measures what it is supposed to measure (Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 1999). Math CBMs have mixed results for content validity. Several studies of math CBMs use the *Monitoring Basic Skills*

Progress measures, a description in which sampling from a state level curriculum is offered in the manual (Fuchs Hamlett, & Fuchs, 1998). One concern that has been raised through a number of studies have discussed that reading demands are inherent in the understanding of mathematics, thus affecting its validity (Christ et al., 2008). The criterion-related validity studies show that coefficients between math CBMs generally range from .26-.67 with the *Metropolitan Achievement Test* (MAT) Operations, MAT Problem Solving, and *District Criterion Referenced* Test Basic Math Concepts (Christ et al., 2008; Marston, 1989). When examining the relationship between math CBMs to commercial achievement tests such as Math Computation and the Concept of Number subtest of the Stanford Achievement Test and the Woodcock Johnson- R Applied Problem subtest, the median coefficients ranged from .74-.83 (Christ et al., 2008).

In summary, CBMs are a useful tool in many areas of student improvement. It allows educators to see if a student is making progress or not making progress, to eliminate unnecessary referrals, and develop a baseline for goal setting for student expectations (Deno, 1986). More recently, as CBMs have become more commonplace in classrooms, attention has be drawn to the fact that CBMs can improve academic accountability and can also be a predictor of performance on high stakes standardized achievement testing (Deno, 2003; Good, Simmons & Kameenui, 2001; Muyskens & Marston, 2002).

Accountability

The idea behind educational accountability has been a concern since the 1980s, when the National Commission on Excellence in Education was established to examine the quality of the education system in the United States. The focus behind the movement was to create standards for what students should know, and use that information to measure where students are performing and create resources for teachers to improve student achievement to meet these standards (Braden & Tayrose, 2008). In 1983, a report entitled, “A Nation at Risk,” described general academic underachievement of the country (Nation at Risk, 1983). Areas included the 40-50 point drop in SAT scores and the inability for 17-year-old students to write a persuasive essay or solve a multistep math problem. In response to this, the commission made 38 recommendations across the areas of Content, Standards and Expectations, Time, Teaching, Leadership, and Fiscal Support. In regards to content, the report recommended that schools have set requirements for the number of years that students take in the areas of English, mathematics, science, social studies, language and computer science. Specifically to math, it is recommended that students take at least three years of mathematics. There were also recommended changes in standardized testing, and extended school days to address the advancement of student achievement (A Nation at Risk).

Prior to this movement, there were no academic standards or expectations for student performance. Teachers were often the targets for blame on students’ academic struggles, but many teachers claimed it unreasonable to hold them accountable when expectations were not clearly defined (Braden & Tayrose). In addition to the development of standards, it was necessary to develop tests to measure if the standards were being attained. Federal laws and policies have mandated the use of state testing and consequences for schools based on the results (Braden & Tayrose).

While standards and a great deal of federal funding have been placed in educational reform, it was not until President George W. Bush’s NCLB Act that accountability was required for student outcomes in general education. The accountability portion of NCLB is the primary idea of the plan to improve public school systems. The law requires states to develop challenging standards in areas of reading and math, administer annual testing to all students in grades 3-8, and to set annual progress goals. NCLB also mandates all states to have 100% of students proficient by 2014 for all groups of students (Braden & Tayrose, 2008).

There is extensive research on reading curriculum-based measures as a predictor of academic success (Deno, 1985). There have been several studies that investigate the relationship between CBMs and statewide-standardized achievement tests, especially in the area of reading. Many researchers have examined the relationships of CBMs in oral reading fluency and performance on statewide assessments in Washington (Stage & Jacobson, 2002), Colorado (Shaw & Shaw, 2002), North Carolina (Barger, 2003), Minnesota (Hintze & Silberglitt, 2005), Michigan (McGlinchey & Hixson, 2004), Florida (Buck & Torgenson, 2003), Illinois (Sibley, Biwer, & Hesch, 2001) and Oregon (Crawford, Tindahl, & Stieber, 2001).

These studies examined the correlation between performance on state testing and oral reading fluency measures for both 3rd and 4th graders. The range of correlations was between .44 (Washington) and .79 (Illinois). On average, most studies reported correlations in the .60 to .75 range (Shapiro et al., 2006). Correlations above .60 are considered to be a strong correlation, demonstrating an acceptable relationship between the two variables (Salvia & Ysseldyke, 2001). Taking into account the range of different states and the fact that each state created their own assessment, there is remarkable consistency between student performance on a reading passage and high stakes testing.

Therefore, this demonstrates that there is a potential powerful link between CBMs and statewide assessment measures (Shapiro et al., 2006).

The search for technically and theoretically appropriate measures, largely resolved in CBM reading, still remains a concern in the area of mathematics. For over 20 years, there has yet to be a consensus on what to measure, or how to measure it, which were questions posed by Deno and Fuchs (1987).

Math Curriculum

The area of math is divided into different subcategories, which subsequently makes curriculum development difficult due to the diverse skills that students are required to learn throughout their elementary school education. Generally speaking, students have shown growth over the past few years of mathematics testing however; there is still great improvement needed. This section will focus on the general subcategories of mathematics, where students are performing and how this impacts their future.

National Math Curriculum Expectations

The area of math is divided into different subcategories such as number sense, computation, concepts and applications, algebra, geometry, calculus, etc. This paper will focus on two broad constructs: computation and concepts/application. Computation is based on practice to develop automaticity in addition, subtraction, multiplication, and division facts (National Math Panel, 2008). Concepts and applications are also known as a problem-solving math. It includes number concepts, numeration, applied computation, geometry, measurement, charts, graphs, and word problems (Fuchs, Fuchs, & Courey, 2005).

Mathematics curricula are created by sequencing types of mathematical skills students are to learn with and across respective grade levels. Different skills are addressed: geometry, measurement, and problem solving. Therefore, there is great diversity in the types of problems students need to learn and subsequently be assessed on (Foegen, Jiban, & Deno, 2007). In reading, the goal is less diverse: students need to learn how to read. Once the initial skills of translating text into oral language are acquired, growth is measured through fluency. While there is some variability such as vocabulary, background knowledge, and comprehension, the nature of reading remains the same. This is why students "learn to read" in the first three grades and then "read to learn" in the upper elementary through high school years (Foegen, Jiban, & Deno, 2007). Similar ideas are not present in mathematics. While students may have mastered simple addition at the earlier ages, concepts that use addition (e.g., algebra, geometry, calculus, etc.) rely on the foundational skills, but the abstract complexity of the operations should be treated like new areas of knowledge (Foegen, Jiban, & Deno, 2007).

In order to assist in promoting national mathematical competency, the National Math Panel (NMP) reviewed research on effective mathematical practices and made general summaries and recommendations in a NMP report published in 2008. The general underlying feature of this report is to improve students' algebraic thinking. There is a great deal of discussion in the report that algebra is the "gateway" to adulthood mathematical proficiency. The panel suggests that our current system "is broken and must be fixed" (NMP, 2008, p.xiii). The panel agrees the important place to start is to put "first things first" meaning, build a strong foundation of math for the nation's youth including conceptual understanding, procedural fluency, and automatic (i.e., quick and effortless) recall of facts (NMP, 2008). Research reviewed by the National Math Panel discusses fluency in computation. Computational facility with whole number operations rests on the automatic recall of addition and related subtraction facts, and of multiplication and related division facts. It requires fluency with the standard algorithms for addition, subtraction, multiplication, and division (NMP, 2008). Few curricula in the United States provide sufficient practice to ensure fast and efficient solving of basic fact combinations and execution of the standard algorithms (NMP, 2008). For all content areas, practice allows students to achieve automaticity of basic skills—the fast, accurate, and effortless processing of content information—which frees up working memory for more complex aspects of problem solving (NMP, 2008). The National Math Panel has developed a few benchmark guidelines for fluency with whole numbers. The National Math Panel suggests that by the end of grade three, students should be proficient with addition and subtraction of whole numbers. By the end of grade five, students should be proficient with multiplication and division of whole numbers (NMP, 2008).

The research is mixed on which is a better predictor of mathematic achievement. According to the National Math Panel (2008), the debate over the importance of conceptual knowledge: the understanding of mathematical concepts such as understanding mathematical relationships, procedural skills (i.e., the understanding of math rules and procedures), and computational math (i.e., automaticity of arithmetic equations) are misguided. Each of these skills is connected to each other and builds upon each other. Understanding of basic mathematics computation and fluency, promote effective and efficient problem solving (National Math Panel, 2008). A recent review of state standards found that 46 of 50 states created standards within mathematics that aligned with the National Council of Teachers of Mathematics, which calls for an enhanced "capability to think and reason mathematically (NCTM, 2000, p. 30) rather than a "curriculum focused on arithmetic procedures" (Burrill, 1997, p.335). The disagreement of mathematical priorities between the NMP and the NCTM highlight the importance of research to determine what should be included in mathematics curriculum in order to promote mathematical proficiency in students.

Current Mathematics Performance

The current mathematics performance of elementary students may be progressing, but there is still great improvement to be made. The National Assessment of Educational Performance and Department of Education (NAEP) gathers data on the national performance of 4th and 8th graders in school subject areas. In 2007, the NAEP assessed students in grades 4 and 8 performance in the area of mathematics. The NAEP test performance results fall into four categories: below basic, basic, proficient, and advanced. Results show that only 39% of 4th graders are scoring at or above the proficient level on mathematics achievement testing and only 35% of 8th graders score proficient or above (Lee, Grigg, & Dion, 2007). As mentioned previously, there have been improvements over time. In 1990, 13% of 4th grade students scored in the proficient range. In 1996, that number rose to 21% and 2003, the numbers rose again to 32% proficient.

Wisconsin performance, grade 4

As a state, Wisconsin 4th grade students tend to perform slightly higher than the national average of public schools. In 2007, the average scaled score in Wisconsin was 244, while the national average was 239 (out of 500 possible points). Of the 52 states & jurisdictions that participated in the assessment, 4th grade Wisconsin students scores were higher than those in 31 jurisdictions, not significantly different than 16, and lower than those in 4 jurisdictions. This shows that Wisconsin 4^{th} grades students are performing relatively well compared to other states. When looking at the percentages of basic, proficient, and advanced levels of performance, in 2007, 85% of students performed at or above the basic level, and 47% at or above the proficient level (Lee, Grigg, & Dion, 2007). Therefore, only half of the 4^{th} grade students are performing at or above the proficient level, which is higher compared to most states, but may not be considered acceptable to the NCLB standards. Similar to reading, NCLB calls for the 100% proficiency in mathematics by 2014 (NCLB, 2001).

The Need for Mathematical Skills

The jobs most in demand in Wisconsin both currently and predicted in the next 10 years will require Science, Technology, Engineering, and Math (STEM) skills and knowledge. STEM jobs dominate the fastest-growing job opportunities and are higher-paying positions (Dickman, Schwabe, Schmidt, & Henken 2009).

Many companies have a minimum mathematics proficiency rate, even if the job does not typically require such skills. Once employed, individuals who are proficient in mathematics earn 38% more than those who are not (Riley, 1997). Due to the need for mathematics proficiency, there is a demand to examine how current students are performing and how performance can be improved (Clarke & Shinn, 2004).

Proficiency in mathematics also is important for individuals, because it gives them college and career options. A solid foundation in high school mathematics through Algebra II or higher correlates with access to college, graduation from college, and earning in the top quartile of income from employment (National Math Panel, 2008). The value of such preparation promises to be even greater in the future. The National Science Board, as stated in the National Math Panel Report, indicates that the growth of jobs in the mathematics-intensive science and engineering workforce is outpacing overall job growth by 3:1 (National Math Panel, 2008). Due to the increase of importance on technology, the national workforce of future years will have to handle mathematical concepts more completely and more skillfully than in today's workforce. Therefore, math is a national issue (NMP, 2008).

Curriculum Based Measurement in Improving and Predicting State Standardized Achievement in Mathematics

The literature on mathematics CBMs and predicting standardized achievement scores is not as extensive as the literature in the area of reading. While probes in mathematics were established during the development of CBMs in the late 1970s-early 1980s, reading has taken the spotlight of most of the research (Shinn, 1989).

Math CBMs and high stakes testing

Research on math CBMs as an indicator on future high stakes testing is sparse and inconclusive. Helwig, Anderson, & Tindal (2002) examined 8th grade students and the effectiveness of a computerized math CBM prediction on an Oregon state achievement test. In this study, CBM math probes predicted with 87% accuracy the students who would meet state math standards (Helwig et al., 2002). While this article shows great promise for math CBMs, it does not account for students in elementary school, which is the crucial age for early identification regarding mathematical difficulty.

Shapiro and colleagues (2006) researched CBMs and performance on state assessment standardized tests on math performance in Pennsylvania. This study examined students in grades three through five. The researchers found correlations consistently in the .50 range when examining the relationship between math computation and the PSSA (Pennsylvania System of School Assessment). When examining math concepts/applications, these researchers found correlations exceeding .60. These correlations are considered to be moderate to strong when examining a relationship between a math CBM and performance on the PSSA. The limitations to the study include the limitation of schools only in Pennsylvania and the issue of attrition, which tends to be a common limitation in any study collecting data over time.

Jiban and Deno (2007) conducted a study comparing computation of basic math facts to performance on a math section of the Minnesota state achievement test in grades three and five. The study found that in grades three, the basic math facts had a correlation from .11-.44, which is vastly different from the correlation found in the study conducted by Shapiro et al. (2006). This level of correlation is considered to be low to moderate at best, meaning there is not a particularly strong relationship between basic math facts performance and performance on the state standardized test. However, it is important to note there was an issue in reliability with the sample size in grades 3; therefore, the authors encourage readers to interpret the results with caution. The sample size for grade three was 35 students, which is considered by the researchers to be an inadequate size in correlational research. In grade five, the reliability was stable and the correlations to the state testing ranged from .55-.60, which are a bit stronger, indicating a stronger relationship between math facts performance and performance on state standardized testing. As mentioned, the authors note that there was an inadequate sample size; therefore additional research needs to be conducted in order to establish the strength of predictive ability of math CBMs on state standardized tests.

Computation vs. Problem Solving: What to Assess?

There has been a debate on which area of mathematics should be the focus of instruction and measurement of proficiency. Early research in mathematics CBMs used basic computation only and included single digit by single digit math fact problems using all four arithmetic operations (Jiban & Deno, 2007). As noted in the previous section, the NMP suggests that today's educational system needs to bolster automaticity of fluency in computational problems (NMP, 2008). Additionally, a student would not be able to complete a conceptual mathematical problem without a strong foundation in computational mathematics (Helwig, Anderson, & Tindal, 2002).

On the other hand, while computation is important, the majority of state standards require that students at all grade levels have a conceptual knowledge of mathematics (Helwig, Anderson & Tindal, 2002). Helwig, Anderson & Tindal argue that conceptual CBMs may have a stronger ability to predict performance on mathematics achievement testing. Conceptual understanding on mathematics is often considered a prerequisite to the successful application of procedures and operations in problem solving situations (Helwig, Anderson, & Tindal; Boaler, 1998, Moss & Case, 1999). There is currently an increasing emphasis on problem solving applications on statewide testing, in which computation alone is not adequate (Helwig, Anderson, & Tindal). It is also important to note that conceptual math involves graphs, shapes, and words, which also relate to the ability to read words, and therefore is not a pure measurement of math (Jiban & Deno, 2007). Because the research is not clear on which form of mathematics serves as a stronger predictor of competency of statewide testing, the current study will examine both computational and conceptual mathematics.

Summary

This information presented allows us to see that as a nation, and as individual states, our students are not performing in mathematics up to stakeholder expectations. Currently, students and schools are assessed by standardized testing that occurs once a year and results are given at the end of a school year, leaving little time for instructional changes for the students that need it. This has many consequences for students, teachers, and administration. There have been several studies of predicting state testing performance using reading CBMs and some promising preliminary studies in using math CBMs in predicting state performance as well. Therefore, CBMs may be a tool that can assist in increasing performance in the area of mathematics. By establishing where students are performing currently and how they measure up to criterion or norm based benchmarks, it is anticipated to predict approximately how they will perform on standardized testing.

This study will look specifically how students in the areas of Wisconsin perform on both computation and concepts and application probes of mathematics and compare them to their performance on the WKCE, Wisconsin's state standardized test. Specifically, grades three through five will be examined and with this information, normative benchmarks will be established along with the strength of the relationship and predictive ability of math CBMs and state standardized test performance.

Purpose and Significance of the Study

The purpose of the current study is to examine the predictive ability of math CBMs, both computational and applied mathematics, on outcomes of the Wisconsin Knowledge and Concepts Exam-Mathematics (WKCE-Math). The current available literature is limited in the area of computational and conceptual math CBMs and its predictive ability to standardized testing. This study will make an addition to the literature and extend the findings to evaluate a potential tool to assist both faculty and students in promoting success in terms of proficiency on mathematics in the classroom as well as on state standardized achievement testing. This study anticipates determining a relationship between CBMs and achievement testing. This information will add to the reliability and validity of CBMs as an efficient tool to gauge mathematical performance.

Research Questions

R_1: How strong will curriculum based measurements in mathematics relate to student performance on WCKE testing?

R_2: In the prediction of WKCE performance, will there be a difference in the relative strength of the relationship of computational math and applied math skills in relation to the grade level of the students?

CHAPTER II

METHODOLOGY

Two critical features of learning mathematics are computation skills and concepts/applications. Both of these skills are not only evident in everyday grade-level mathematics curricula, but also are frequently found in state standardized testing. These skills build upon each other and it is essential to have solid computation skills in order to progress to higher levels of mathematical thinking, such as concepts and applications (NMP, 2008).

Participants

The current study used 97 elementary students from one school. There were 35 third grade students, 25 fourth grade students, and 37 fifth grade students. All students participated in the process. Students in grades three through five were studied because this is the age that elementary students take the WKCE. Additionally, the school district initiated the study to improve the math instruction in the school.

Procedures

The school district initiated the collection of data for math CBMs and WKCE. Therefore this is an archival study of the resulting data with no identifying information. To obtain current performance on math computation and concepts and applications, I administered the MBSP probes as part of my practicum expectations in the school.

Instructions were followed precisely from the technical manual in order to promote consistency and reliability.

Data was collected during the second week of the school year, then two weeks later, and a final probe was administered one week prior to taking the WKCE in grades three through five. The WKCE was administered in late October. Due to the need to compare test score of the CBMs and WKCEs, it was necessary to have students to provide their names on all measures; however, all student information was converted to a number to be entered into the data base and correlated with the performance on the WKCE. The information will remain confidential and be disposed of in accordance with the American Psychological Association (APA) guidelines.

Materials

The students were given math probes from the *Monitoring Basic Skills Progress-Basic Math Computation* and *Monitoring Basic Skills-Concepts/Applications*. Fuchs, Hamlett, and Fuchs developed both probe booklets in 1998. Both probes have established reliability and validity that is listed in detail below. Additionally, both have been used in previous research regarding math CBMs (Foegen, Jiban, & Deno, 2007; Shapiro et al., 2006).

Computation

Monitoring Basic Skills Progress-Math Computation (2nd edition) probes consisted of one sheet of paper consisting of 25 problems of addition, subtraction, multiplication, and division based upon the grade level. As the grades advanced, the skills required increase in difficulty. Each grade level has 30 alternative forms of probes. The probes used in this study were randomly selected for each grade. Students in third and fourth grade had 3 minutes to complete the probe and students and fifth grade had five minutes to complete the probe. This was consistent with instructions provided in the technical manual. An example of a 3^{rd} grade computation problem found on the probe is "5x4". An example of a 4^{th} grade problem is "3/7-2/7", and an example of a 5^{th} grade problem is 2.8-1.58.

The technical manual reports two studies of alternate form reliability that examine both single scores and aggregations of two scores (Fuchs, Hamlet, & Fuchs, 1998) The first study sampled 79 students with mild disabilities in grades one through six, single form reliability by grade level ranged from r = .73 to r = .92, with aggregation improving reliabilities to r = .91 to r = .96. The second study, sampled 48 students without disabilities in grades one through six, single form reliability, ranged from r = .83 to r = .93, with aggregated score reliabilities, ranging from r = .93 to r = .99.

Criterion validity was studied using 65 students with mild disabilities. Mean scores from multiple CBM forms were correlated with those on the Math Computation Test (Fuchs, Fuchs, Hamlett, & Stecker, 1991), the Stanford Achievement Test (SAT) Concepts of Number subtest, and the SAT Math Computation subtest. When separated by the grade level which was assigned for appropriate monitoring, these correlations ranged from r = .49 to r = .93; when all student scores were treated as a unified group, correlations with criteria ranged from r = .66 to r = .83.

Math Concepts and Applications

The probes in Monitoring Basic Skills Progress-Math Concepts and Applications were used to assess students' ability in math concepts and applications. There are 24 problems for grades 3-5 which are designed to measure their knowledge of grade level concepts and applications. These probes cover measurement, number concepts, charts and graphs, money, fractions, applied computation, and word problems. Some problems required a simple, one part answer, while other problems required three answers. Grades 3 and 4 had six minutes to complete the probe and grade 5 had seven minutes to complete it. Each part of a problem answered correctly was given a point. The total amount of points across the probe was used to measure performance. An example of a 3rd grade application problem is using the ruler provided on the page, measure the crayon to the nearest inch. An example of a 4th grade problem is 1 hour= ___minutes. An example of a 5th grade problem is finding the average of these numbers: 19, 7, 12, 8, 9.

Internal consistency over time was measured for the Concepts and Applications probe. The mean score from all odd-numbered measures was correlated with the mean score from even-numbered measures, each mean constituting an aggregation of 10 to 15 scores. These correlations, separated by grade level, ranged from .94 to .98.

Criterion validity was studied using the same sample, with the Comprehensive Test of Basic Skills (CTBS)-Computation, -Concepts and Applications, and -Total Math scores serving as criteria. Correlations between a mean of students' last three CBM scores and the criteria ranged from .66 to .81. Correlations between the MBSP Concepts and Applications scores and the MBSP Computation scores ranged from .63 to .90. Thus there was acceptable validity for the measure.

WKCE

The WKCE is a standardized measure given to all students in grades 3-8 and grade 10 enrolled in Wisconsin public schools. The WKCE was developed by McGraw

Hill, under the direction of the Department of Public Instruction (DPI). The WKCE measures areas of reading/language arts, mathematics, science, and social studies. The scope of this paper will focus only on the mathematics portion. The math test consists of two types of responses: selected-response (multiple-choice) or constructed-response format. The math test design takes about 80% of the total score points from selected response items and 20% of the score points from student-generated constructed-response items. The WKCE scoring breaks down performance into proficiency categories: advanced, proficient, basic, and minimal performance. The "advanced" category demonstrates in-depth understanding of academic knowledge and skills tested on WKCE at that grade level, "proficient" demonstrates competency in the academic knowledge and skills tested on WKCE at that grade level, "basic" demonstrates some academic knowledge and skills tested on WKCE at that grade level, and "minimal" performance demonstrates very limited academic knowledge and skills tested on WKCE at that grade level. Wisconsin strives for proficient performance. The mathematics portion measures six objectives: mathematical processes, number operations and relationships, geometry, measurement, statistics and probability, and algebraic relationships.

Reliability was assessed in a number of ways. When a constructed response to a mathematical operation was evaluated, inter-rater reliability was between .79 to 1.00. Other questions only involved the selection of a response. For these questions, the internal consistency coefficient, represented by Cronbach's apha, for the mathematics subtest ranged from .91 to .93 across the test and various groups (e.g., minority status, grade levels, etc.) (Wisconsin Department of Public Instruction, 2008).

Validity was assessed through content and construct validity. Content validity was established through a process of gathering expert teachers and developing a content blueprint of the test based on the Wisconsin Model Academic Standards. Focus groups were periodically commissioned by the Department of Public Instruction (DPI) to work with the test developer to ensure that appropriate questions that addressed the Model Academic Standards are retained. Other reviewers address potential bias, curricula alignment, and presentation variables. DPI had the final authority for the content of the test (Wisconsin DPI, 2008). For construct validity, differential item functioning was used to ensure there was no potential bias across the various groups (e.g., male/female, minority status, etc.) based on group membership. In addition, correlations between standards assessed by the math subtests reveled that there were moderate correlations (i.e., .55 to .66) between the subscales indicating that each strand measures a similar but unique area of math skills (Wisconsin DPI).

Hypotheses and Data Analysis

> HO_1: There is no difference in the performance on the math CBMs across grade levels.
>
> H_1: There is a difference in the performance on the math CBMs across grade levels.

In order to determine if there are differences in grades three through five, a one-way MANOVA is necessary. The independent variable is the grade level and the dependent variable is the math probe (computation vs. concepts/application).

> HO_2: There will be no relationship between CBMs and WKCE for third grade, fourth grade, and fifth grade students.
>
> H_2: There will be a relationship between CBMs and WKCE for third grade, fourth grade, and fifth grade students.

This question addresses the relationship between the CBM math probes (computation and concepts/application) and the WKCE. In order to do this, it is necessary to carry out a multiple regression for each grade level independently with the WKCE as the independent variable and the math probes as the dependent variable.

CHAPTER III

RESULTS

This study is designed to examine the relationship between curriculum-based measurement in mathematics and WKCE scores. It is hypothesized that WKCE scores can be used to predict future performance on state standardized assessments, specifically, the WKCE. The study examined two different probes, math computation and math concepts and application, to see if there was a difference in the probes, the grade level, and the relationship to the WKCE.

Descriptive Statistics

A simple frequency count was run to determine the amount of males and females (see Table 1) that were present in each grade level and the number of students who scored in each WKCE category (see Table 2).

Table 1. Descriptive Statistics for Gender

	3^{rd} Grade	4^{th} Grade	5^{th} Grade	Percentage
Males	19	16	15	49%
Females	19	10	22	51%

Table 2. WKCE Performance Across Grades 3-5

	3^{rd} grade	4^{th} grade	5^{th} grade	Percentage
Minimal	9	0	3	11%
Basic	6	2	6	14%
Proficient	12	13	17	42%
Advanced	11	11	11	33%

Internal Consistency Reliability

Cronbach's Alpha was used to determine the reliability within probes. Reliability for third grade and the computation probe ranged from .78 to .86. In fourth grade, reliability between computation probes ranged from .62 to .71 and in fifth grade, the reliability between computation probes in the fifth grade ranged from .55 to .70. Third grade concepts/application probes demonstrated internal consistency reliability from .75 to .77. Fourth grade concepts/application probes ranged from .52 to .67, and fifth grade probes ranged from .65 to .75. Most of the reliability between probes ranged between

.60’s and .70’s. A score of .60 and above is considered to be adequate (Salvia, Ysseldyke, & Bolt, 2007). See Table 3 for results for each specific probe.

Table 3. Internal Consistency Reliability of Math CBM probes

	Comp A	Comp B	Comp C	App A	App B	App C
3rd	.78	.82	.86	.76	.77	.75
4th	.62	.71	.66	.67	.52	.62
5th	.55	.70	.62	.69	.65	.75

Average Performance on Probes

Students were given three probes, two weeks apart, to assess for test-retest reliability, A, B, and C respectively. The average scores on the probes tended to increase in all three grade levels for both computation and concepts/applications. A one-way MANOVA was conducted with grade level as the independent variables and type of math problems as the dependent variables (i.e., computation and application). The overall MANOVA test was significant, $F(4, 188) = 27.65$, $p < .001$. Scores on the computation were significant across grade level, $F(2, 95) = 8.64$, $p < .001$. A Shefffe post hoc analysis revealed that third grade student performed significantly different from fourth and fifth grade students. For the third grade computation probe, performance ranged from 7.5 to 12.3 with the median across three administrations being 10.1 digits correct. For the 4th grade computation probe, performance ranged from 12.2 to 22.6 with the median being 18.4. The 5th grade computation scores ranged from 13.0 to 16.9 with the median score 15.2. Performance on the concepts application were also significantly different by grade level, $F(2, 95) = 22.67$, $p < .001$. A Scheffe post hoc analysis revealed a significant difference across all three grade levels. For third grade concepts/applications the problems correct ranged from 13.2 to 15.1 and the median score was 15.1. For fourth grade concepts/applications, performance ranged from 10.3 to 14.3 with the median being 12.2. For 5^{th} grade the problems correct ranged from 6.9 to 9.0 with the median being 7.7. See Figure 1 and 2 for comparison of performance on probes.

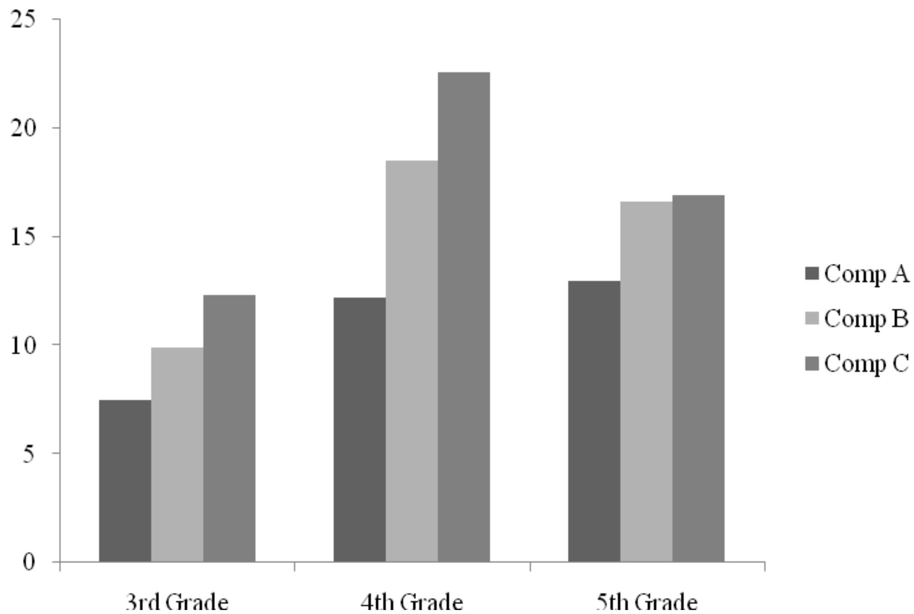

Figure 1. Average Digits Correct for Math Computation Probes by Grade

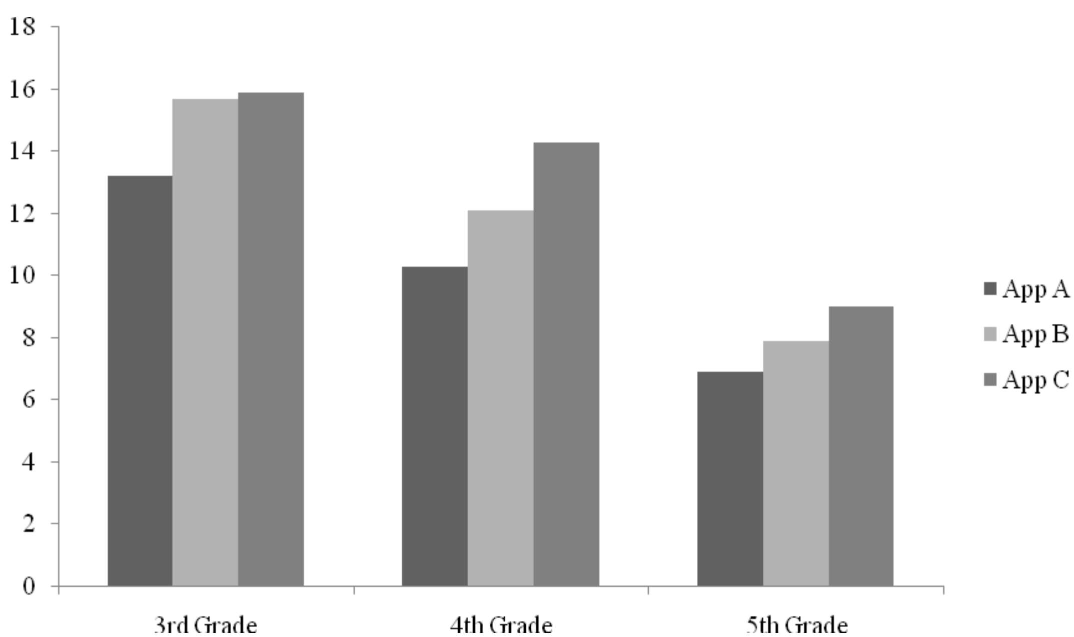

Figure 2. Average Problems Correct for Math Application Probes by Grade

Correlations by Grade

Correlations were determined for each grade for each of the median probes, and for the WKCE scores. All correlations for the third grade probes were statistically significant with each other, ranging from .58-.79 when correlated with WKCE scores. All 4th grade probes demonstrated a statistically significant relationship with each other, and .54-.63 when correlated with the WKCE. Fifth grade probe correlations showed a statistically significant correlation with each other and correlations ranging .45-.65 when correlated with the WKCE.

Third Grade

The first probe, Comp A, correlated highly to the other computation probes, Comp B and C (.81 and .74 respectively). Additionally Comp A was also correlated with the application probes App A, B, and C with correlations of .72, .69, and .64 respectively. Similarly the other probes did correlate with each other and to the alternative probe (computation or application). In third grade, correlations for the probes and the WKCE came out fairly moderate to high, ranging from .58 to .79. All of these were found to be statistically significant ($p < .05$). The strongest correlations to the WKCE were Computation B and Application B with correlations of .74 and .79. The median scores for computation and application probes also correlated highly with the WCKE with correlations of .70 and .73 respectively. See Table 4 in Appendix for 3rd grade correlations.

Fourth Grade

For the 4^{th} grade data, the first computation probe, Computation A, correlated with the other computation probes, B and C, with results of .80 and .76 respectively. Application A correlated with Application probes B and C with correlations of .82 and .81. Additionally, Computation probes correlated to Application probes with correlations ranging from .32 to .81. All but one correlation was considered to be statistically significant (Computation A to Application A=.32). All probes demonstrated a statistically significant relationship ($p<.05$) to the WKCE with correlations ranging from .54 to .63, with the strongest correlations to the WKCE being Computation C (.63) and Application B (.60). The median scores for the computation and concepts/application probe correlated adequately with the WKCE, with correlations of .62 and .57 respectively. See Appendix table 5.

Fifth Grade

For 5^{th} grade data, Computation A probe correlated with B and C with correlations of .64 and .61. All computation probes in fifth grade demonstrated a statistically significant relationship when correlated with each other. Additionally, Application A correlated with Application probes B and C at the statistically significant level with correlations of .76 and .63 respectively. Similarly, all application probes demonstrated a statistically significant relationship when correlated to each other. Computation probes also demonstrated a statistically significant relationship when compared to application probes with correlations ranging from .38 to .69. Both probes also demonstrates statistically significant relationships to the WCKE scores ranging from

.56 to .65, with the strongest correlations being probes Computation B (.65) and Application A, B, and C (all with correlations of .65). The median computation and concepts/applications correlated strongly with the WKCE with correlations of .61 and .64. See Appendix Table 6.

Regression Analyses

Overall Regression Analysis

An overall analysis of the computation and application probes across the three grades levels was conducted to predict the WKCE Math score. Grade level, computation probes and application probes all significantly predicted WKCE performance, $R^2 = .70$, F(4,93 = 54.63, p < .001, 95% CI [.61, .79].

A series of stepwise regressions were completed for each of the grade levels. Pairs of probes and significant results can be found in Table 7-9. In third grade, the best predictor of WKCE performance was the second probe set, Computation B and Application B . Application probes tended to enter first in the data set. See Table 7.

Table 7. Stepwise Regression Predictors of WKCE Math Scores for Third Grade Students

	WKCE Math Score		
		Model 2	
Variable	Model 1 *B*	*B*	95% CI
Constant	346.02	352.67	[328.24, 371.73]
Median App.	5.09	3.36	[1.36, 5.36]
Median Comp.	--	1.92	[0.43, 3.41]
R^2	.54	.61	
F	41.97*	27.80*	
ΔR^2		.08	
ΔF		6.84	

Note. $N = 38$. CI = Confidence Interval.

* $p < .001$

In fourth grade, the best predictor of WKCE performance was computation. When the median probe set was entered into the equation, only computation was significant. Application scores did not add to the power to predict WKCE scores. See Table 8.

Table 8. Stepwise Regression Predictors of WKCE Math Scores for Fourth Grade Students

	WKCE Math Score	
	Model 1	
Variable	*B*	95% CI
Constant	433.75	[406.13, 461.38]
Median Computation	2.53	[1.15, 3.90]
R^2	.38	
F	14.37*	

Note. *N* = 26. CI = Confidence Interval.

* *p* = .001

In fifth grade, the best pair in predicting WKCE scores was the first set, Computation A and Application A. The application only probes tended to result in more significant values without the addition of the computation probe. The median computation and application probe were significant in predicting WKCE scores with the Application probe entering the equation first. See Table 9.

Table 9. Stepwise Regression Predictors of WKCE Math Scores for Fifth Grade Students

	WKCE Math Score		
		Model 2	
Variable	Model 1 *B*	*B*	95% CI
Constant	437.00	428.35	[405.30, 451.41]
Median App.	6.44	4.27	[1.05, 7.48]
Median Comp.	--	1.67	[0.13, 3.20]
R^2	.41	.48	
F	24.41*	15.98*	
ΔR^2		.07	
ΔF		4.85	

Note. N = 37. CI = Confidence Interval.

* $p < .001$

Overall, both computation and application probes were significant in predicting WKCE results. The 3rd grade equation accounted for the most variance. CBM Math probes need to contain computation and application components at 3rd grade. At 4th grade, computation produces more significant results. Finally, for 5th grade, both application and computation seem to be important. The predictive power of the 4th and 5th grade probes on important outcomes, while significant, tended to be lower than the third grade data set.

CHAPTER IV DISCUSSION

OF RESULTS

This study examined the relationship between math computation and application probes and their predictive validity to the mathematics portion of the Wisconsin Knowledge and Concepts Examination (WKCE). Overall, the purpose of the study was to determine whether the Math CBMs were a predictor of the WKCE and to closely examine computation and application skills to see if one more closely aligns with the WKCE than the other.

The 3rd grade probes demonstrated the strongest predictive ability for WKCE results. It was hypothesized that computation probes would have more of an influence on WKCE scores than application probes; however, this was not the case. Both computation and application played an important role in predicting WKCE results with concepts/applications loading higher than computation. Both computation and application account for 61% of the variance. This would indicate that both computation and concepts/application probes could be useful tools in predicting WKCE performance. The results of this study are similar to those in Shapiro et al (2006), who found similar magnitudes of correlation between computation and concepts and application in grades 3-5. Jiban & Deno (2007) found much lower correlations between math computation CBMs and state testing. However, they did not factor in math concepts and applications.

The consistency of the current study's results along with previous research in regards to the magnitude of the correlation between math probes and standardized testing tells us there is a strong relationship between the results of math CBM probes and scores on the state standardized testing.

At 4th grade it appears that computation is still a very important factor in predicting math outcomes. This is contrary to the initial hypothesis that applied problem solving will have a greater predictive ability than computation. Results from the 4^{th} grade data show that only computation loads in the predictive validity. However, the significance of the application from loading on the equation was .08 and the cut for significance was .05. Given the lower numbers for this population, it would be important to keep application with computation for the 4th grade probes. This outcome may have been impacted by several different factors. When looking at the 4^{th} grade data set f or computation, it appears that in general, 4^{th} grade students had a much higher average digits correct than 3^{rd} and 5^{th} grade. This may be a result of several factors. The students may have simply been guessing correctly on the probes. It is more likely that the data may be a reflection of the curriculum being used in school as the probes may reflect what was being taught in class at the time. Another possibility may relate to the general ability of the cohort that may be above the general ability of grades 3 and 5.

Although the scores on the probes tended to be higher, generally, the correlations between the 4^{th} grade probes and the WKCE scores were lower than those of 3^{rd} and 5^{th} grade. This may be a result of the smaller sample size, as 4^{th} grade had about 10 fewer overall students than 3^{rd} and 5^{th} grade. In addition, it may be a reflection of the probe and its relationship to the WKCE mathematics portion. While students may be taught pencil/paper computation in class and succeed in the area, the WKCE may place stronger emphasis on reading skills.

The results in this study in regards to the magnitude of correlation were somewhat lower than correlations found in earlier research. The most concerning information found in the stepwise regression is the variance in which the probes can predict WKCE outcomes, which is less than 50%, meaning it is more likely that other factors influenced the predictive ability than the relationship itself. With this information, schools should likely proceed with caution in using these particular probes in the school for the 4^{th} grade class, unless more data is collected over time proving greater predictive ability and ruling out outside factors such as cohort ability, guessing, and other conditions.

In 5^{th} grade, application did take on more importance in predicting math outcomes. This confirmed the original hypothesis that math concepts and applications were stronger predictors of WKCE performance. I hypothesized that 5^{th} grade students would have mastered numerical operations because they are more prevalent in early elementary school curriculum. As a result, mathematical concepts and applications would be more prevalent in both the 5^{th} grade curriculum and the WKCE mathematics portion because these skills require a higher level of mathematical problem solving. Due to the higher magnitude of the correlations and the predictive ability than found in prior research (Jiban & Deno, 2007; Shapiro et al., 2006), both concepts/applications and computation probes used in conjunction can be fairly accurate tools to measure 5^{th} grade student achievement as it relates to WKCE performance.

One major finding is related to the time between administrations of the probes. For each grade level, the average score on the probes had increased for each administration. This shows that there was potentially too much growth in the two-week time span between administrations of probes. This information could be a result of practice effects or developmental growth. In addition, this could be a result of the curriculum, especially with the impending WKCE exam; teachers could have been placing extra emphasis on building the skills of computation and concepts/application, therefore showing a rapid increase in performance across all grade levels.

Generally speaking, math CBMs can be a valid predictor of WKCE performance in grades 3-5. The probes need to measure both computation and application for the best predictive ability, thus, making it a useful tool for teachers and school psychologists to measure a student's performance before they take the standardized test.

Limitations

There were some variables that may have impacted the results of the research. As a result of using data for one rural school, the sample size is small. While the results from this study seemed to be similar to other studies measuring math CBMs to state standardized testing, a larger sample size may have impacted the final results, especially in regards to the significance of the correlations and predictive ability. This may be particularly true in 4^{th} grade, as that population had the smallest sample size and the weakest of the three correlations. As a result, the probes used in 4^{th} grade may not be as reliable and should be used with caution.

Future Research

The data collected and analyzed for the purpose of this study provided support for the predictive ability of math CBMs on WKCE performance. However, it is important to continue collection of data on these students and continue to compare their performance to future years' WKCE performance. This would allow the school track student progress and performance over time, allowing for more data to use in support of instructional changes and early intervention to promote proficiency on state assessments.

As mentioned in the literature review, research in the area of math curriculum based measurements and predicting future performance on assessments is limited. Therefore, any research in this area would contribute to what has been done in the past. It would be beneficial to have a greater sample size with a greater amount of diversity to see if different populations would have different results.

Recent information put out by *Our Nation's Report Card (2007)* not only suggests that the students of the United States are struggling, but there is a gap within the students based on race and income. White, Black, Hispanic, and Asian/Pacific Islander students all showed higher average mathematics scores in 2007 than in any of the past assessments of mathematics. However, score increases did not show a significant closing of performance gaps between minority students and White students. There was no significant change in the White – Black score gap over the last two years. The White – Hispanic gap was not significantly different from the gaps in 2005 (Lee, Grigg, & Dion, 2007). Therefore, research conducted involving a greater diversity in student population and their performance on math CBMs and WKCE could provide useful information to assist in continuing to close the education gap.

Socioeconomic status is factor that can impact a student's performance on state testing. In *Our Nation's Report Card* (2007), a student's eligibility for free or reduced-price school lunch is used to classify socioeconomic status; students from low-income families are typically eligible, while students from higher income families typically are not (Lee, Grigg, & Dion, 2007). Students from higher socioeconomic status continued to score higher on average than students who were eligible for free or reduced-price lunch. This information should be examined in future studies, examining whether or not these probes could be useful in early identification in students of various race and ethnicities, along with socioeconomic status, as schools are continuing to become more and more diverse in population.

Reading was not addressed in this current study, however, it may be important to examine in future research. Reading plays a part in students' ability to perform concepts/application problems. If students struggle with reading the word problems, they would likely experience difficulty in solving those particular problems. While this would not likely impact computation, it would be interesting to see how WKCE reading scores or other measurement of reading correlated with the mathematics probes.

Future research should also consider changing the time frame for administration of the probes. As noted in earlier in the research paper, CBMs can have several uses such as a progress monitoring tool or a tool to collect baseline data. In this study, the average score on the probes had increased for each grade level. This demonstrates that the probes used in the study are sensitive to growth, which is crucial for a progress monitoring tool. However, two weeks appeared to be too long between administrations using the probes as a baseline assessment tool, and subsequently may have impacted the data collected. In the future, this probe should be used more rapidly, perhaps across a few days rather than a few weeks.

Implications for School Psychology

Research on reading CBMs have demonstrated their usefulness in examining a student's performance and using that data to make decisions for curriculum changes as well as educational placement. As noted previously, the information for math CBMs is not as conclusive. The data collected through this study allows school psychologists to see that math CBMs can be useful classroom tools to not only see a student's progress or struggles in computation and concepts/application mathematics, but also use the data to inform curriculum changes so students are more likely to improve not only their overall academic performance, but specifically on the WKCE mathematics section.

Impact on Educators

This information not only impacts the field of school psychology, but also impacts the field of education, as these probes are implemented in the classroom. Subsequently, it is crucial for classroom teachers to understand the purpose of the assessment as well as its interpretation. The CBM probes are a formative assessment and allow classroom teachers to 'gauge' student performance. This allows for immediate feedback received by teachers and therefore they can make more frequent changes to

their own instructional practices. This will be difficult for teachers to accomplish if they do not understand the data they are collecting. As a result, school psychologists and teachers need to work side by side to promote complete understanding of the assessments, data, and the interpretation.

RtI and Early Intervention

The results of this also results of this study also contribute towards the support of the RtI model and its practice in school psychology. These CBMs can be used essentially for the screening of students in the general education setting, which is considered to be Tier I of interventions. By administering CBMs and using norms, teachers can identify which students are in need of intervention immediately, whereas WKCE results do not come in until months after the test is taken. Math CBMs provide immediate feedback, which allows for immediate interventions to help the student reach greater mathematics success.

As a result of early intervention, not only may students have increased performance on WKCEs or state standardized testing, but when looking at the big picture, may have prevented unnecessary psychoeducational testing or inappropriate educational placement. By using CBMs to determine the effectiveness of the instruction of the three tiers of RtI, educators can modify the instruction based on the need of the student (Reschly), rather than immediately refer a student for a special education assessment. This is crucial to promoting the paradigm shift from “refer-test-place” to RtI.

Conclusion

As a nation, students are not meeting stakeholder expectations in the area of mathematics. Currently, students and schools are assessed by standardized testing that occurs once a year. The results are given at the end of a school year which leaves little time for intervention for the students that experience difficulties. There have been some promising preliminary studies in using math CBMs in predicting state performance; however a majority of the research in this area has focused on reading. The goal of this research study was to determine the practicality of computational and concepts math

CBMs to predict performance in grades 3-5 on a statewide assessment in mathematics. The research indicated that generally speaking, math CBMs can be a valid predictor of WKCE performance in grades 3-5. The probes need to measure both computation and application for the best predictive ability, thus, making it a useful tool for teachers and school psychologists to measure a student's performance before they take the standardized test.

REFERENCES

American Educational Research Association. (2000). *AERA Position Statement on High-Stakes Testing in Pre-K – 12 Education*. Washington, D.C.

Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville: North Carolina Teache r Academy.

Boaler, J. (1998). Open and close mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29,* 41-62.

Braden, J. P. & Tayrose, M. P. (2008). Best practices for school psychologists in educational accountability: High states testing and educational reform. In Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (Vol 2, pp. 575-588). Bethesda, MD: National Association of School Psychologists.

Bradley-Johnson, S., & Dean, V. J. (2000). Role change for school psychology: The challenge continues in the new millennium. *Psychology in the Schools, 37,* 1-5.

Buck, J., & Torgeson, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report 1). Tallahasse: Florida Center for Reading Research.

Burrill, G. (1997). The NCTM standards: Eight years later. *School Science and Mathematics, 97,* 335-339.

Chambers, J, Parish, T, & Harr, J. United States Department of Education, Special Education Expenditure Project. (2004). *What are we Spending on special education services 1999-2000?* . Washington, DC: U.S. Department of Education, Office of Special Education.

Christ, T. J., Scullun, S., Tolbize, A., & Jiban, C. (2008). Implications of recent research: Curriculum-based measurement of math computation, *Assessment for effective intervention, 33,* 198-205.

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.

Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*, 303-323.

Dickman, A., Schwabe, A., Schmidt, J., & Henken, R. (2009). Preparing the future workforce: Science, technology, engineering and math (STEM) policy in K12 education in Wisconsin. Public Policy Forum.

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

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

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

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

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

Erikson, R., Ysseldyke, J., Thurlow, M., & Elliot, J. (1998). Inclusive assessments and accountability systems: Tools of the trade in educational reform. *Teaching Exceptional Children, 31*, 4-9.

Espin, C.A., & Fogen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content area tasks. *Exceptional Children, 62*, 496-514.

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

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488-499.

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

Fuchs, L.S., Hamlett, C. L., & Fuchs, D. R. (1998). Monitoring basic skills progress: Basic math computation (2nd ed.). Austin, TX: Pro-ed.

Fuchs, D., Fernstrom, P., Reeder, P., Bowers, J., & Gilman, S. (1992). Vaulting barriers to mainstreaming with curriculum-based measurement and transenvironmental programming. *Preventing School Failure, 36*, 34-39.

Fuchs, L. S., Fuchs, D., & Courey, S. J. (2005). Curriculum-based measurement of mathematics competence: From computation to concepts and applications to real-life problem solving. *Assessment for Effective Intervention, 30*, 33-45.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*, 518-537.

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

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disabilities Quarterly, 25,* 33-45.

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

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

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. *Journal of Special Education, 36,* 102-112.

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

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

Jiban, C. L., & Deno, S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance: Are one-minute measures technically adequate? *Assessment for Effective Intervention, 32*, 78-89.

Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999)

Standards for Educational and Psychological Testing. American Educational Research Association. Washington, DC.

Kelley, B., Hosp, J. L., & Howell, K. W. (2008). Curriculum-based evaluation and math: An overview. *Assessment for Effective Intervention, 33,* 250-256.

Lee, J., Grigg, W., and Dion, G. (2007).*The Nation's Report Card: Mathematics 2007* (NCES 2007–494). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*, 4-16.

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

McLane, K. (2001). *Curriculum Based Measurement and Statewide Tests*. National Center on Student Progress Monitoring. Washington, D.C: National Center on Student Progress Monitoring.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.*), Curriculum-based measurement: Assessing special children* (pp 18-78). New York, NY: The Guilford Press.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18,* 109-117.

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

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education*, *30,* 122-147.

Muyskens, P. & Marston, D. B. (2002). *Predicting success on the Minnesota Basic Skills Test in reading using CBM.* Unpublished manuscript, Minneapolis Public Schools.

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

National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC.

No Child Left Behind Act of 2001., Pub. L. No 107-110, H. R. 1, 115 Stat 1425 (2002, January 8).

Pollock, J. E. (2007). *Improving student learning one teacher at a time*. Alexandria, VA: Association for Supervision and Curriculum Development.

Powell-Smith, K. A. (2004). *Individual differences in FCAT performance: A national context for our results.* Paper presented at the annual meeting of the Pacific Coast Research Conference, Coronado, CA.

Reschley, D. J. (2008). School psychology paradigm shift and beyond. In Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (Vol 1, pp. 3-16). Bethesda, MD: National Association of School Psychologists.

Riley, R. W. (1997). *Mathematics equals opportunity*. District of Columbia, U. S.: Federal Department of Education. (ERIC Document Reproduction Service No. ED 415119).

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston: Houghton Mifflin.

Salvia, J., Ysseldyke, J. E., Bolt, S. (2007). Assessment in Special and Inclusive Education, 10th ed. New York: Houghton Mifflin.

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

Shaw, R., & Shaw, D. (2002). DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP) (Technical Report). Eugene: University of Oregon Press.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York, NY: The Guilford Press.

Shinn, M. R.. (2008). Best practices in using curriculum-based measurement in a problem solving model. In Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (Vol 2, pp. 575-588). Bethesda, MD: National Association of School Psychologists.

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 achievement. Paper presented at the Annual Meeting of the National Association of School Psychologists. Washington, DC.

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

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

The National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. Washington, D.C.: Government Printing Office.

Tindal, G., Germann, G., & Deno, S. (1983). Descriptive research on the Pine County norms: A compilation of findings (Research Report No. 109) Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement*. University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G. (1992). Evaluating instructional programs using curriculum-based measurement. *Preventing School Failure, 36*, 39-44.

Wisconsin Department of Public Instruction. (2008). *Wisconsin knowledge and concepts examination: Fall 2007 technical manual*. Monterey, California: CTB/MCGRAW-HILL.

APPENDIX A

CORRELATIONS FOR GRADES 3-5

Table 4. Correlations for 3rd grade

	1	2	3	4	5	6	7	8	9
1. Comp A	-	.81	.74	.72	.69	.64	.86	.72	.58
2. Comp B		-	.88	.58	.69	.60	.98	.68	.74
3. Comp C			-	.58	.67	.69	.92	.70	.68
4. App A				-	.82	.77	.62	.86	.64
5. App B					-	.86	.69	.95	.79
6. App C						-	.60	.95	.67
7. Comp Med							-	.67	.70
8. App Med								-	.73
9. WKCE									-

Table 5. Correlations for 4th grade

	1	2	3	4	5	6	7	8	9
1. Comp A	-	.80	.76	.32	.43	.34	.82	.37	.54
2. Comp B		-	.84	.49	.50	.46	.99	.53	.60
3. Comp C			-	.53	.55	.56	.87	.58	.63
4. App A				-	.82	.81	.49	.89	.55
5. App B					-	.79	.52	.96	.60
6. App C						-	.46	.90	.56
7. Comp Med							-	.54	.62
8. App Med								-	.57
9. WKCE									-

Table 6. Correlations for 5th grade

	1	2	3	4	5	6	7	8	9
1. Comp A	-	.64	.61	.46	.44	.41	.77	.45	.56
2. Comp B		-	.73	.38	.56	.55	.90	.57	.65
3. Comp C			-	.45	.66	.69	.90	.64	.63
4. App A				-	.76	.63	.48	.87	.65
5. App B					-	.77	.62	.96	.65
6. App C						-	.62	.84	.65
7. Comp Med							-	.62	.61
8. App Med								-	.64
9. WKCE									-