2011

The Statistical Relationship Among Number Sense, Computational Fluency, and Montana Comprehensive Assessment System

Amy Jo Burland
The University of Montana

Statistical Relationship Among Number Sense, Computational Fluency and Montana Comprehensive Assessment System

by
Amy Burland
B.S., University of Minnesota Moorhead, Moorhead, MN, 1983
M.Ed., University of North Dakota, Grand Forks, ND, 1992

Dissertation

Presented in partial fulfillment of the requirements
for the degree of

Doctor of Education
in Curriculum and Instruction

The University of Montana
Missoula, MT

May 2011

J.B. Alexander Ross, Associate Dean of the Graduate School

Dr. Trent Atkins, Committee Chair
Department of Curriculum and Instruction

Dr. Georgia Cobbs, Committee Member
Department of Curriculum and Instruction

Dr. David Erickson, Committee Member
Department of Curriculum and Instruction

Dr. Ke Norman, Committee Member
Department of Mathematical Sciences

Dr. Darrell Stolle, Committee Member
Department of Curriculum and Instruction

ABSTRACT

Burland, Amy, Ed.D., Spring 2011 *Education*

The Statistical Relationship Among Number Sense, Computational Fluency, and the Montana Comprehensive Assessment System.

Chairperson: Dr. Trent Atkins

According to recent studies, less than half of U.S. students perform at the proficient or advanced levels in mathematics by the time they reach grade 4 and the trend continues through high school. In order to improve instruction many districts have adopted scientifically based researched programs such as Response to Intervention (RTI), which allows for the examination of the effectiveness of the core curriculum that is being used in a school or classroom. In addition, RTI provides school administrators and teachers with educational tools to identify students who may be at-risk of failing and to inform teachers of supplemental instruction needed to build up skills that are identified as weak or lacking. Research on early mathematics skills indicates that skills performance at the kindergarten and first grade level may predict performance at later grade levels. Providing intervention early has been shown to have a positive effect on students' future mathematics success.

This study investigated the long-term predictive validity of the AIMSweb measures for kindergarten through grade 2 and the Montana Comprehensive Assessment System (MontCAS). The kindergarten and grade 1 assessments included the Test of Early Numeracy which measures number sense skills that include Oral Counting, Number Identification, Quantity Discrimination, and Missing Number. The grade 1 and 2 Mathematics-Curriculum Based Measures assessed computational fluency. The scores on these K-2 assessments were analyzed to investigate correlations with the grade 3 MontCAS scores of the same students. The results indicated that Number Identification and Quantity Discrimination provided the most explained variance. Overall, the kindergarten scores were stronger indicators of grade 3 performance than the grade 1 scores.

A sequential multiple regression model was also used to explore which of the TEN measures along with the hierarchy of tests from kindergarten through grade 2 had the greatest explained variance for the grade 3 MontCAS. The results showed that each test from kindergarten to grade 1 increased the predictability of the grade 3 MontCAS scores; however, the grade 2 scores did not contribute to the predictability of the grade 3 assessment. Overall, Oral Counting indicated the highest explained variance using this model.

Acknowledgements

It is a pleasure to thank those that made this dissertation possible beginning with Dr. Trent Atkins, Committee Chair, for his support and guidance throughout the process beginning with choosing my topic and problem to the final writing stages. In addition, I want to thank my Committee Members Dr. Ke Norman, Dr. David Erickson, Dr. Georgia Cobbs, and Dr. Darrell Stolle for challenging me in the way I thought about and approached the field of education and for forcing me to stretch my thinking and gain insight to the overall process of research. They led me to learn each stage of the work from collecting data to analyzing it in more ways than I could have ever thought of on my own. In doing so, my study became more and more interesting and surprising by what can be observed from a small amount of data. I want them to know how appreciative I am of the time they gave to me by meeting, reading, and helping me refine my work till it was completed.

I have special thanks for my Salish Kootenai College Department Chair, Dr. Cindy O'Dell, and my colleagues for making this possible to finalize with the support and assistance at work. Dr. O'Dell has been an inspiration and support in making sure I was able to juggle my work responsibilities with the additional focus on completing coursework and the final stages of this dissertation. Special thanks goes to my patient mentor/teacher/tutor/statistics guru Dr. Merle Farrier who generously gave of his time to help me proceed with the direction given by my Committee.

Last but not least I want to thank my husband, David, and son, Dominic, for their support and patience throughout this journey. Their understanding and belief in me always made me know that it was possible to make it to the completion of this undertaking. I also want to thank my mother and sisters for their encouragement and blessings that I always felt from miles away.

Table of Contents

List of Tables

List of Figures

Chapter One

Background of the Problem

The development of mathematical understanding has come to be known as the "new civil right" in the United States. With only 37% of U.S. fourth graders and 34% of eighth graders performing at proficient and advanced levels, mathematical achievement and success have become a social justice interest defined as the right of all students to be given the opportunity to receive high quality mathematics instruction (Lago & DiPerna, 2010; National Assessment of Educational Progress [NAEP], 2009; Thurber, Shinn, & Smokowski, 2002). This is based on the fact that today's students who develop mathematical understanding are granted far more opportunities and career choices in shaping their future (Boaler, 2008; Jordan, Glutting, Ramineni, & Watkins, 2010; Lago & DiPerna, 2010).

Advancements in technology and information have produced an urgency to train more people in mathematics to work in the fields of medicine, science, and technology. In a broader sense, advancements in these fields have made understanding mathematics necessary for all students rather than for a select few (Boaler, 2008; Jordan et al., 2010; Lago & DiPerna, 2010; National Council of Teachers of Mathematics [NCTM], 2000). For example, most if not all workplaces have been impacted by the increased emphasis on technology and require all employees to come with higher skill levels in mathematics than ever before (Lago & DiPerna, 2010). The same can be said about how technology has changed daily life, where fundamental mathematics language is used to describe national budgets, profits, inflation, demographics, and global warming (Boaler, 2008).

In the past decade the most notable federal action taken to improve K-12 mathematics instruction includes the 2001 reenactment of the Elementary and Secondary Educational Act

(ESEA) entitled No Child Left Behind Act (NCLB). This bi-partisan statute was passed in an attempt to make schools accountable in providing quality education that allows all students to be academically successful, as measured by student performance on federally approved assessments, student attendance, student graduation rates, and teacher credentials. Schools that do not meet Adequate Yearly Progress (AYP) based on federal goals are labeled as needing school improvement, corrective action, or restructuring; therefore, these schools are mandated to take measures that improve the system through comprehensive school reform. Action to improve must be implemented by using scientifically research-based approaches in curriculum and professional development. In addition, these schools are required to provide families with quality services outside of the school program at district expense until they meet AYP for two consecutive years (U.S. Department of Education, 2002).

Response to intervention (RTI). Traditionally, the special education screening process was initiated by a "wait to fail" and then an intelligence quotient (IQ) discrepancy assessment. The reauthorization of the Individuals with Disabilities Education Act (IDEA) in 2004 reconstructed this process because it did not allow for inadequate instruction nor did it allow for the fact that kindergarten and primary elementary students have not had enough academic instruction to accrue an IQ discrepancy. The reconstruction of the screening has mainly seen the implementation of the Response to Intervention (RTI) approach, which is an evidence-based system where students progress through increasingly intensive levels of a prevention process. After progressing through the three-tiered levels, only students who do not show achievement through standard instruction are formally assessed for special education services (Fuchs, 2004; Seethaler & Fuchs, 2010). In addition, with the increased emphasis on schools to demonstrate accountability and meet AYP under NCLB, early identification of students at-risk of struggling in mathematics in later grades and intervention for identified students has taken center stage. Even though RTI has been used in education settings for decades, for the above reasons it has recently come back into educational methods as a “new” approach to identify learning disabilities and/or students who are at-risk of failing. Many states adopted this process because it meets the NCLB standards of an approach that is based on scientific research and provides a structured intervention protocol (Kashi, 2008). According to Kashi (2008), it has been used especially in districts with minority and English-as-a-Second-Language (ESL) student populations and/or other cases where the district students are underperforming academically. The structure of RTI has been proven to effectively individualize instruction regardless of inherent diversity of the overall classroom’s cognitive differences (Kashi, 2008).

This first chapter provides background on the need for early identification and intervention for students who are at-risk of low-performance in K-12 mathematics, as well as an overview of the AIMSweb Mathematics-Curriculum Based Measure (M-CBM) tools that are used to monitor and guide instruction for students entering elementary school. These educational tools include the Test of Early Numeracy (TEN) for kindergarten and first grade students and the M-CBM for grades 1 through 8. The current study compared the TEN and M-CBM scores to grade 3 Montana Comprehensive Assessment System (MontCAS) scores in order to investigate correlations among tests as students progress from kindergarten to third grade. Accordingly, this longitudinal study reports findings from analysis of the predictive validity and reliability of TEN and how it relates to M-CBM and performance on third grade MontCAS. This overview provides insight into the purpose and outline of the current study.

Number sense. The development of Number Sense begins pre-verbally at infancy through experiences that relate number to spatial representations. For example, the toys that infants are given to play with represent sets of small objects, which are precise representations of larger objects (VanDerHeyden, 2010). In fact, infants have a natural number sense that builds through their observations of the world around them, as they categorize and organize information to make sense of it. Such experiences provide a foundation for acquiring symbolic mathematics understanding, which is built up by the indirect use of the following: base ten number system, counting, sorting, comparing, and using number operations. An understanding of verbal and symbolic number systems depends upon early experience and can be successfully taught in preschool and kindergarten (Jordan, Kaplan, Ramineni, & Locuniak, 2009). Thus the importance of early identification and intervention during preschool years sets forth the experience and mathematics skill development in K-12 years (Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett, 2005; Jordan et al., 2010).

Noted researchers have not produced a concise clear definition of number sense because no two researchers define it in the exact same way (Berch, 2005; Gersten & Jordan, 2005). Berch (2005) compiled a list of number sense features while studying number sense literature in the areas of mathematical cognition, cognitive development, and mathematics education. He found "that number sense reputedly constitutes an awareness, intuition, recognition, knowledge, skill, ability, desire, feel, expectation, process, conceptual structure, or mental number line" (p. 333). These numerical features allow one to be flexible in understanding simple to complex mathematical procedures. These understandings range from whole number quantities, such as how adding and subtracting relate to quantities, to making mathematical connections with mathematical operations and procedures, to inventing mathematical procedures, to recognizing numerical errors, and finally to being able to communicate mathematically (Berch, 2005).

Although operational definitions vary for number sense, there are commonalities in the tasks associated with measuring number sense in preschool and kindergarten children. These tasks include quantity discrimination, counting objects, counting aloud, number identification, basic computation, estimation, understanding measurement concepts, number production, and identifying a missing number (Bryant & Bryant, 2008; Lago & DiPerna, 2010). Using these early indicators to identify children who may experience difficulties in understanding mathematics later allows for early interventions at the most receptive periods of a child's education (Lago & DiPerna, 2010). Control group comparison studies show that teaching early number competencies to kindergarteners results in significant gains in first grade mathematics skills (Jordan et al., 2009).

For the above reasons, a child's level of number sense in preschool and kindergarten puts him/her on a path of either mathematical success or mathematical struggles throughout elementary and beyond (Gersten & Chard, 1999; Jordan et al., 2010; VanDerHeyden, 2010). Early intervention can make a difference since children at this level are more likely to retain and learn information. Furthermore, early intervention can prevent emotional and behavioral problems that contribute to repeated mathematics failure and frustration in later years (Lago & DiPerna, 2010).

Test of Early Numeracy (TEN). The TEN is one of the tools used for the current study. This test is made up of four individual one-minute assessments administered to kindergarten and first grade students. The short administration time makes them repeatable throughout the year, age appropriate, and nonintrusive in the classroom instructional setting. They are conducted as a natural part of the curriculum because they are specific to the skills that build a foundation for understanding symbolic mathematics (Clarke & Shinn, 2004). Again, one of the purposes of the

TEN/M-CBM design is to assist teachers in identifying children at an early age who are at risk of failing to understand mathematics as they progress through elementary, middle school, and high school.

Mathematics curriculum-based measurement (M-CBM). The AIMSweb TEN and M-CBM system is an RTI component, which involves individual formative computation assessments that are conducted in the fall, winter, and spring for each student. These assessments inform instruction by identifying areas of difficulty in basic skills. The assessments narrowly encompass one concept at a time and track growth and mastery from which teachers design interventions that give the student a higher likelihood of future success (Christ & Vining, 2006; Keller-Margulis, Shapiro, & Hintze, 2008).

Early identification. Gersten, Jordan, and Flojo (2005) studied kindergarten, first, and second grade children's performance in mathematics. They concluded that some students were "typical" in performance and understanding while others showed difficulties in basic fact computation, counting strategies, and *number sense,* which they define as a strong understanding of number relationships and basic understandings of the size of numbers. Gersten et al. (2005) found that as students move on to intermediate and upper elementary school lacking the understanding of basic facts, they continue to have difficulties throughout their education. Such mathematical difficulties are broad-based in regard to the misunderstandings of concepts and behaviors exhibited in applying mathematics concepts. These difficulties are shown by students of all ages (Bryant & Bryant, 2008; Jordan et al., 2010; Lago & DiPerna, 2010).

In the current study, predictive validity was investigated on kindergarten and first grade students using the EM-CBM or TEN. This system measures number sense through four specific assessments that include oral counting (OC), number identification (NI), quantity discrimination

(QD), and missing number (MN). The OC measure requires the participant to begin with one and orally count for one minute. NI requires students to use a list of random numbers ranging from 1 to 20 and to orally identify them. The QD measure requires the participant to look at two numbers and to name the number that is larger. Lastly, the MN measure requires the participant to identify the missing number in a three-number string of consecutive numbers within the 1 to 20 range (Clarke & Shinn, 2004).

The formative assessment used to investigate predictive validity for grades 2 through grade 3 was the AIMSweb M-CBM. This assessment is considered curriculum-based because it is aligned with the grade level curriculum and emphasis on expected computational skills for each grade level. The system provides 40 alternative forms for each grade level use for Benchmark Assessment, Strategic Monitoring, and frequent Progress Monitoring. Again, this assessment made it possible to test the interventions being used in a time efficient way that is sensitive to change over time (Shinn, 2004).

Traditional versus standards-based reform mathematics instruction. Computation is the hallmark or fundamental mathematical foundation for students in grades 1 through 6. This is the central agreement between the *traditional* and the *standards-based reform* approaches to mathematical instruction. This seems to be the only agreement between the two polarized approaches when discussing the pedagogy behind each (Baroody, Bajwa, & Eiland, 2009).

The traditional approach is teacher-centered—meaning that the teacher teaches students in a linear fashion addressing a single concept at a time. This approach has the students solve each problem by using memorized algorithms given by the teacher or textbook followed by as many as 75 practice problems. The traditional approach stresses basic fact rapid recall and is usually assessed through repetition and timed tests (Boaler, 2008).

Traditional methods are said to teach by working through an operation one place value at a time with transition to an adjacent position (trades or regrouping) thinking in terms of digits rather than the composite number that the digits make up (Kamii, 2000). This approach is criticized for teaching students to use memorized formulas and prescribed strategies—omitting the opportunity for them to make conjectures and strategize ways that helps them make sense out of concepts. Without this experience students become more concerned with restating the formulas and steps dictated to them than understanding mathematics and building confidence in their own constructs (Boaler, 2008).

On the opposite end of the instructional spectrum is the *standards-based reform* approach which was introduced in the late 1980s and reinforced through the NCTM Principles and Standards (2000). This approach uses discovery and inquiry based methods that promote higher order thinking, reasoning, and problem solving. A majority of mathematics instruction time is spent by students interacting with peers to find solution strategies for problems that integrate a range of concepts (e.g. geometry, statistics, probability, measurement, and number sense) (Bryant & Bryant, 2008). It requires primary students to solve real-life problems through discussion and investigation as well as through games that prompt them to use various number operations while discovering the relationships among the operations and properties of numbers. Children are required to invent their own flexible strategies through an emphasis on understanding whether their solution makes sense based on their experiences and mathematical knowledge. The reform approach does not suggest teaching specific traditional strategies until grades 5 or 6 when students have had ample experience with the base ten system and have had time to construct their own knowledge through these discovery experiences. By inventing their own strategies and observing peers develop different ways to solve problems, students acquire the foundation for higher mathematic skills based on mastery of operations and number properties (Russell, 2000).

Criticism of the standards-based reform approach appears in research findings that focus on children with mathematics learning disabilities. These studies suggest that the reform approach in and of itself is insufficient for struggling students. According to Bryant & Bryant (2008), these students need explicit strategic instruction that teaches subskills and uses a combination of procedural rules, metacognitive cues, memory retention and retrieval techniques, and mnemonics. Small group pullout sessions work best so that students receive immediate feedback and take part in more interaction with peers and teacher. This approach is an intervention service component of RTI (Bryant & Bryant, 2008).

Purpose of the Study

The purpose of this study was to investigate the long-term predictive validity of TEN measures that include oral counting (OC), number identification (NI), missing numbers (MN), and quantity discrimination (QD). These dynamic indicators were analyzed to learn more about their validity as an EM-CBM. This study analyzed the correlations among student TEN scores to student mathematics scores on grades 1 and 2 M-CBM, and grade 3 MontCAS. Originally, an analysis of grade 4 scores was included but not enough data were available. In addition, the study was also to analyze concurrent validity with grade 3 M-CBM and grade 3 MontCAS scores but again, there was not enough data available. Correlations for each of the TEN probes were analyzed individually. The research questions for this study were as follows.

Research Questions

1. Did an Oral Counting test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?

2. Did a Number Identification test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?
3. Did a Missing Number test of early numeracy in kindergarten and grade 1 correlate with mathemathics performance in grade 3?
4. Did a Quantity Discrimination test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?
5. Did Mathematics-Curriculum-Based Measurement in grades 1, 2, and 3 predict student performance on third and fourth grade MontCAS?
6. Which of the Test of Early Numeracy (TEN) measures explained the most variance on the MontCAS and M-CBM assessments?

$H_{01:}$ There would be no statistical significance in the relationship between the Oral Counting test of early numeracy in kindergarten and grade 1 and proficiency in grade 3.

$H_{02:}$ There would be no statistical significance in the relationship between the Number Identification test of early numeracy in kindergarten and grade 1 and proficiency in grade 3.

$H_{03:}$ There would be no statistical significance in the relationship between the Missing Number test of early numeracy in kindergarten and grade 1 and proficiency in grade 3.

$H_{04:}$ There would be no statistical significance in the relationship between the Quantity Discrimination test of early numeracy in kindergarten and grade 1 and proficiency in grade 3.

$H_{05:}$ There would be no statistical significance in the relationship between M-CBM and the end of year Montana Comprehensive Assessment System (MontCAS) for grades 3 and 4.

$H_{06:}$ Each of the Test of Early Numeracy (TEN) measures would equally explain variance on the MontCAS.

Definition of terms

Adequate Yearly Progress (AYP): A mandate under the 2001 reenactment of the Elementary and Secondary Education Act (ESEA) entitled No Child Left Behind (NCLB) which states that all public elementary and secondary school students must meet academic achievement standards and schools must work toward narrowing achievement gaps among different socio-economic, racial, and gender groups. Measures include performance on standardized tests, attendance rate, graduation/drop-out rates, retention rate, and percentage of students completing advanced placement and gifted programs (U.S. Department of Education, 2002).

Computational Fluency: Ability to use basic mathematics facts and foundational skills that include calculating with flexibility, efficiency, and accuracy (Boerst & Schielack, 2003).

Curriculum Based Measurement (CBM): The standardized formative assessment procedure that monitors student academic growth in fundamental skills relevant to school outcomes (Christ & Schanding, 2007; Shinn & Bamonto, 1998).

Early Mathematics Curriculum Based Measurement (EM-CBM): Assessment tools used to identify kindergarten and first grade students who are at-risk of failing in mathematics during later grades (Clarke & Shinn, 2004).

High Stakes Tests: State and national standardized assessments that measure student proficiency with consequences imposed on schools that fail to show student success; therefore, "high stakes" pressure on administrators, teachers, and students impact educational systems through funding and renewal of teacher and administrative contracts (Reys & Lappan, 2007; United States Department of Education, 2004).

Mathematics Curriculum Based Measurement (M-CBM): An established procedure of formative assessment that measures individual student growth and mastery of computational fluency. The assessments are aligned with grade level curriculum standards and materials (Christ & Schanding, 2007; Keller-Margulis et al., 2008).

Mathematics Proficiency: Federal measures, which apply uniform annual measurable objectives in mathematics to all students across the United States and territories. Student mathematics mean scores on standardized tests are used to measure how well a school is meeting the objectives and AYP (Kim & Sunderman, 2005).

Number Sense: The understanding of whole number quantities and relating addition and subtraction to quantity while making connections that provide a foundation to the understanding of symbolic mathematics (Jordan et al., 2010). This mathematical understanding allows for flexible ways to solve problems through construction and deconstruction of numbers (Gurganus, 2004; Jiban & Deno, 2007; NCTM, 2000).

Response to Intervention (RTI): An intervention approach based on scientific research that provides continuous monitoring of individual student data. Student performance objectives designed to meet individual student needs are used to show growth (Kashi, 2008).

Summary

This study was intended to provide information for primary educators who seek early intervention techniques for students who are at risk of failing in mathematics. Early intervention is pivotal in providing quality instruction that allows students to progress at standardized levels throughout the K-12 curriculum. Analyzing individual skills provides more information for teachers, as it points out which early number skills are most critical to predicting success on third and fourth grade MontCAS tests, which directly impact AYP status. This same information can be shared with parents to promote early mathematics experiences for preschoolers within their daily lives.

Educators who use this study and others like it will find that expanding the use of formative tests can be tools for demonstrating accountability to state and federal agencies. In fact, formative assessments that measure teaching through student learning may be better than summative standardized assessments as an indicator of high quality instruction that prepares students for deeper understanding and higher performance in mathematics. Ideally, using this system as an accountability measure would make it possible to eliminate high stakes tests and at the same time increase instruction time lost to the current mandated testing system.

Chapter 2

Review of Literature

The importance of mathematics scores has become more and more pronounced in political and social sectors throughout the world, including the United States. At a time when global economics has more and more emphasis, U.S. students' mathematics test scores rank well below their peers in other advanced countries. Indeed, it is each student's right to be provided with a quality mathematics education to ensure that he/she truly has a future in choices of lifestyle and career. For these reasons, research on quality instruction and intervention at the earliest educational level is important to provide future opportunities for all students.

This chapter examines investigative mathematics assessment research approaches that provide intervention for students who are at-risk of failing to understand mathematics at the intermediate and higher level of education. In addition, Chapter 2 first summarizes the current challenges that schools and teachers face in the political and social arenas. The chapter also reviews historical and theoretical frameworks for RTI approaches that are used for early identification of mathematics difficulties. Third, research on number sense is presented in order to define preschool and kindergarten mathematics skills that are needed for success in higher levels of mathematical skills. Further explanation of the relationship between specific early mathematical skills and contextually relevant variables is provided. The chapter concludes with a discussion of the reliability and validity found in current research on primary grade early mathematics skills.

Social and Political Pressures Facing Schools

Nationally, one in five students receives special education services in order to lessen an achievement deficit. Research on reading disabilities has more attention than mathematics learning disabilities (Thurber et al., 2002). Reading deficit researchers claim that the number of studies conducted regarding reading creates a false assumption that students struggle less with mathematics, but in reality, national mathematics tests show that only 39% of fourth graders and 34% of eighth graders perform at proficient and advanced levels (NAEP, 2009; Thurber et al., 2002). Previous research in mathematics has mainly focused on young children who become learning disabled in mathematics as they progress through elementary levels. More research needs to be done on students at all achievement levels including those who struggle but are not identified as having a cognitive learning disability (Bryant & Bryant, 2008).

Scientifically Research Based (SRB) Programs and Assessments

As was referred to earlier in Chapter 1, the 2001 enactment of the NCLB Act holds schools accountable in providing quality education that allows all students to be academically successful. Success as defined by a school that meets AYP, is measured based on student performance on standardized tests, attendance, and high school graduation rates among other administrative standards including teacher qualifications (United States Department of Education, 2004).

NCLB requires schools that do not meet AYP to institute comprehensive reforms that implement programs that have been shown to be successful through scientifically based research (SBR). The NCLB Act (2001, section 9101) defines SBR as rigorous, systematic, and objective procedures to obtain reliable and valid knowledge. The Act (2001) breaks this definition down further to explain that SBR requires empirical methods that involve observation or experiments that test a stated hypothesis and then justify the conclusions. Furthermore, these studies must be clearly detailed to enable them to be replicated or at least built upon in other studies. Lastly, to be considered SBR, a study must show that it was published by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific process (U.S. Department of Education, 2002).

One SBR approach is RTI, which began use in education settings decades ago but has recently come back into educational methods as a “new” approach to identify learning disabilities. In addition, it has been found to assist district students who are academically underperforming—based on standardized tests used to demonstrate whether or not districts meet AYP and compliance with the NCLB mandate (Kashi, 2008).

History of RTI

Response to Intervention (RTI) can be traced back to B.F. Skinner’s work in the 1950s. It has evolved over decades because the early framework was thought to be too expensive for a regular school district to implement. The charting system alone at different stages required classroom teachers to enroll in a semester-long course to become trained in using the system. Researchers continued to study ways to provide effective instruction for basic skills, building upon the work of educators striving to reform the common educational approaches. The best approaches used in public schools were those that monitored individual student progress; therefore, that was the main focus of the work being studied (Crawford & Ketterlin-Geller, 2008; Kashi, 2008).

During the 1980s and 1990s, more longitudinal individualized monitoring procedures were introduced including the Personalized System of Instruction (PSI). This criterion-referenced system was not widely accepted because it lacked conformity with grade level distributions; however, it still made sense in the way of individualized monitoring of students. For this reason, it kept the interest alive to continue to work toward an individualized system that was realistically cost effective and would improve on the general instruction and assessment for early intervention of academic difficulties in basic skills (Crawford & Ketterlin-Geller, 2008; Kashi, 2008).

Theoretical Framework and Response to Intervention (RTI)

RTI is an evolution of special education protocols that traditionally used a combination of one-point-in-time assessments such as intelligence, achievement, and behavior evaluations to make special education decisions for students at risk of academic failure. This screening process was controversial because it failed to be a valid and reliable decision making process for low-performing students. In addition, it failed to provide ongoing effective intervention that led to positive results for students (Barnett, Daly, Jones, & Lentz, 2004).

In 2002, the President's Commission on Excellence in Special Education echoed the criticism of the traditional assessments saying that they do not provide functional outcomes to make special education decisions. Functional outcomes, they claimed, are those that lead to social and academic trajectories for low-performing students and students with disabilities. The report went so far as to recommend a total abandonment of the traditional classification system and replace it with a decision-making process based on response to instruction. The recommendation required a scientific process shown to be valid and reliable through continuous progress monitoring and do away with the "wait to fail" model previously used in special education identification (Barnett et al., 2004).

The President's Commission on Excellence in Special Education (2002) followed soon after the enactment of NCLB. In 2004, IDEA adopted the Commission's recommendations so that children with disabilities would be served through SBR programs and all students would receive effective instruction and progress monitoring in the general education classroom; especially before entering special education programs (Barnett et al., 2004). The purpose of utilizing progress monitoring over the one-time assessments was to assure that individual student needs were identified and effective interventions were designed to meet those needs as students progressed (Gilbertson, Witt, Duhon, & Dufrene, 2008).

For these reasons, more and more general education teachers are trained in RTI to guide their instruction and make effective educational decisions that meet individual student needs. Schools and districts are using it as an SBR approach for NCLB comprehensive reform efforts in meeting AYP (Kashi, 2008).

Response to Intervention (RTI) Process

RTI in most cases involves a multi-tiered approach aimed at preventing students from experiencing long-term inadequate general classroom instruction that develops into extenuating existing disabilities (Bryant & Bryant, 2008; Stecker, Fuchs, & Fuchs, 2008). The first step of RTI involves training classroom teachers on scientifically validated instruction. Teachers are trained to use the general classroom curriculum data to make instructional decisions that are individual-based. This allows gaps in student learning to be discovered early-on through formative assessments and intervention. Tier two uses the information provided in tier one to provide supplemental support for students who demonstrate needs. Tier three provides intensified instruction specific to needs still unmet which sometimes includes special education services (Hoover & Love, 2010). These actions take place before students are failing to keep up with peers and standard grade level benchmarks. With RTI, students who demonstrate a low level of achievement receive intervention services earlier and not only make gains on specific skills but also build on skills that may have caused them to lag at more progressive stages of the curriculum (Bryant, Bryant, Gersten et al., 2008).

Multi-tiered prevention and intervention approaches have become common for early reading programs but more research needs to be done for use in early mathematics. Research that investigates relationships between specific weaknesses identified in early mathematics understandings and later mathematics difficulties is essential to improving instruction to benefit all students (Bryant, Bryant, Gersten et al., 2008).

Number Sense and Critical Early Mathematics Skills

Number sense is defined by NCTM (2000) as the ability to view numbers in a flexible manner where one can decompose numbers as well as developmental computation strategies using reference points such as 10 or 100 or 1/2. Moreover, it is a progression of mathematical understanding that involves moving from the initial development of standings on the size of numbers, number relationships, patterns, operations, and place value. Number sense is also described as number knowledge or the ability of understanding quantity. This is especially important and significant for first graders so that they develop skills that allow them to begin calculating in their heads and understanding the fundamentals of how a base ten system works in regard to place value (Bryant et al., 2008).

Developing number sense as a young child is critical if students are to avoid mathematical difficulties as they progress through the elementary grades and beyond. It allows them to develop a sense of basic mathematical concepts that lead to mastery and fluency in manipulating arithmetic combinations. In short, students with number sense gain command of numbers instead of developing math anxiety through the belief that numbers command them. Student understanding of how to compose and decompose numbers puts them in control of the numbers and makes them confident in their mathematical abilities. Progressing in this way allows students to gain sophistication in their strategies and allows them to build an understanding of the magnitude and or quantity of numbers (Locuniak & Jordan, 2008). This becomes evident in their representations, their ability to compute mentally, and in their mathematical explanations (Bass, 2003; Boaler, 2008; Kamii, 2000).

Van de Wall et al. (2010) found that an early understanding of the base ten system is a critical part of mathematics instruction because students who do not master the system are challenged in multi-digit operations and calculations. The progression of place-value understandings that children must advance through by the end of first grade includes:

a. Single numeral: where a student sees a number such as 36 and thinks of it as just 36—without cognizance of any other representation.
b. Position names: where a student can name the digit place value positions (i.e. 3 is in the ten's place and 6 is in the one's place) but still no awareness that 36 is not only 3 tens and 6 ones but is also 36 ones or 2 tens and 16 ones and so forth (decomposing and composing numbers).
c. Face value: using place value blocks, a student will place 3 ten-blocks to the left of 6 ones-blocks but still not understand the quantity composition of the number.
d. Transition to place value: students begin to register quantity of 3 tens and 6 ones demonstrated by placing 30 ones-blocks in the one group left of 6 ones-blocks.
e. Full understanding: students group 3 sets of tens-blocks and another 6 blocks for the ones.

Preschool children informally develop skill in working with arithmetic combinations by grouping and partitioning objects. In primary grades these skills are strengthened by opportunities to solve simple problems of basic facts through the use of arithmetic combinations. Students who struggle with arithmetic combinations are identified as early as first grade when they show difficulties in counting strategies, recalling basic facts from long-term memory, and decomposing and composing numbers. Longitudinal studies that follow students with these difficulties in primary grades demonstrate the same weaknesses as they advance in grade level (Gersten & Chard, 1999; Jordan et al, 2010). The difficulties noted above detour them from developing computational fluency.

Computational Fluency

NCTM (2000) defines computational fluency as having efficiency, accuracy, and flexibility. Efficiency requires a student to solve problems using steps that keep him or her working toward a solution without getting lost but always understanding the logic behind each step. Accuracy involves careful recording of steps with an understanding of number combinations and an understanding of how number operations and number properties relate. Flexibility requires the ability to choose an appropriate strategy with which to solve the problem and the ability to double-check it using a second strategy. Flexibility refers to flexible thinking. In short, fluency requires more from students than just memorizing a single procedure. Students who learn procedures before developing a deeper understanding of number operations and properties and how they relate, will depend on memorized rules.

The traditional approach to teaching students involved providing traditional algorithms or rules and is said to have students work through digits (i.e. “carrying” and “borrowing” according to the rules) rather than numbers (i.e. 14 is understood to be 12 + 2 or 10 + 4 or 15 – 1 and so forth.) (Kamii, 2000). Critics of the traditional approach claim that it does not allow students to develop the understanding and command over their ability to decompose and compose numbers (breaking apart or combining) as they design a strategy to solve a problem (Kamii, 2000). An example of a student decomposing a number can be seen with the problem of 57 x 2. The traditional approach uses multiplication moving from right to left where the student takes 2 x 7 and "carries" the 1 then solves for 2 x 5 and adds the 1 to make 114. A student with number sense and a command of how numbers and properties relate would know to think of 50 x 2 as 100 and 7 x 2 as 14 which again computes to 114. Once students build that understanding and command of numbers throughout the primary grades, traditional algorithms and rules make sense (Van De Walle et al., 2010).

This is how students develop long-term computational fluency and in turn build more sophisticated number sense connections. Furthermore, students become more flexible in developing strategies, more efficient in finding solutions, and more accurate in computation (Russell, 2000).

Sometimes computational fluency is thought of only as accuracy (the right answer) and efficiency (done with speed). Without flexibility—a higher level of thinking where a student can recognize whether a solution is reasonable and makes sense—is absent (Bass, 2003). As important as accurate and exact answers are, according to Reys (1998), a deep understanding of numbers requires that students making sense of "what is reasonable" is even more important. This requires a highly developed number sense that allows reflective and higher-order thinking. This is not only a trajectory to the following grade level but serves students throughout their lives (Reys R. E., 1998).

Computational fluency and number sense may be defined separately but research on mathematical knowledge provides strong evidence that they develop simultaneously. According to Griffin (2003), a student who has developed one has always developed the other and vice versa. She gives an example of this based upon an item on the Number Knowledge Test (NKT), which asks questions similar to, "If Tom has 4 marbles and then Mary gives him 3 more marbles, how many marbles will Tom have all together?" There are five levels of performance on the test and the first level is demonstrated by most preschoolers of age 3 or 4 as they show no attempt to solve the problem. Four and five year-olds at the second level will tend to give a solution such as *a lot*, *5,* or *10 marbles.* According to Griffin (2003), students at this second level, responded with at least five, which indicates mental evaluation and developing number sense (not necessarily computation at this point). At 5 years old, the third level is demonstrated by counting from one with their fingers—sometimes without redeeming the exact solution but closer than the previous level and demonstrating computation. Opportunities to solve similar problems contribute to stronger number sense and computational skills while leading to more sophisticated strategies in Level 4. At this level, 5 and 6 year-olds begin to count-on meaning they will start at 4 and count up 3 more numbers to find their solution. Finally at level 5, children will automatically retrieve the answer mentally and demonstrate a long progression of development that occurs over the years. At this level, if a child is asked how they figured it out, they tend to explain that they just knew it. This happens through the progression that develops from lowest to highest levels as both their number sense and computation skills advance.

Intervention for Students Who Lack Number Sense

Interventions for students with arithmetic difficulties should begin with numeracy skills in preschool or kindergarten. Basic understandings of quantity need to be reinforced when students demonstrate a lack of knowledge at the very earliest of stages in kindergarten. First grade interventions need to continue with basic skills in order to get the rudimentary foundation needed for long-term progress in arithmetic achievement (Bryant, Bryant, Gersten et al., 2008).

The intervention process should begin by identifying students with difficulties in the regular classroom setting and assessing specific skills that are lacking. Based on the assessments and observations, interventions within or outside of the regular classroom should be implemented. Such interventions should emphasize building student skills with relationships of 10 and basic fact computational fluency (Bryant, Bryant, Gersten et al., 2008).

The range of interventions for ages 3 to 6 begin with the 1 to 10 number sequence and continue all the way through mentally computing simple problems. A rich environment that provides a variety of repeated opportunities for counting and solving problems eventually leads to inventing and modeling strategies that are progressively more efficient. Simple computation and number sense rich classrooms embed a network of meaning that directly links to computational fluency and number sense (Griffin, 2003). The intervention process provides a foundation for future mathematic achievement (Bryant, Bryant, Gersten et al., 2008; Griffin, 2003).

What is Curriculum Based Measurement (CBM)

The CBM process is a central component of RTI assessment because it establishes an initial baseline for student achievement that is used to measure changes in student performance. Shinn and Bamonto (1998) describe CBM results as dynamic indicators of basic skills—dynamic in the way of being an assessment that is sensitive to student differences. CBM assists teachers to individualize instruction for students by pinpointing each student's understanding of individual basic skills and concepts. In addition, frequent assessment results are tracked over time to measure change and to consistently inform teachers of reinforcements needed by individual students. The scores provide an accurate picture of a student's performance in a broad number of tasks using basic skills (Shinn & Bamonto, 1998). It is used to determine the level and pace that students progress in learning basic skills (Christ & Vining, 2006; Thurber et al., 2002). It allows instructional short-term and long-term goals to be set after a student is placed at the appropriate level.

Formative assessments are needed for intervention so that deficits are identified early and modified instruction is implemented to address the findings. Conducting only summative assessments does not provide opportunity for intervention because they are given after instruction while formative are given during instruction. Indeed, formative evaluation informs teachers of modifications needed for individual students to increase summative assessment performance (Thurber et al., 2002).

The mathematics curriculum based measurement (M-CBM) is frequently conducted as short, quick, and easy computation tests. They are usually two to five minutes in length (Thurber et al., 2002). Two main constructs that are measured in M-CBM include computation and application mathematic skills. Computation refers to use of mathematical operations. Application skills involve solving word problems through knowledge of number patterns, measurement, and operations to devise strategies to solve problems (Thurber et al., 2002).

AIMSweb Assessment

AIMSweb is an M-CBM system and product of the Pearson Company (2010), which provides RTI formative assessments for students in K-8. As was stated earlier, the EM-CBM for kindergarten and early-year first grade students is the TEN, which is made up of four 1-minute assessments that are done with individual students. This assessment measures number sense through four specific tests that include Oral Counting (OC), Number Identification (NI), Quantity Discrimination (QD), and Missing Number (MN). The OC measure requires the participant to begin with 1 and orally count for one minute. NI requires students to use a list of random numbers ranging from 1 to 20 and orally identify them. The QD measure requires the participant to look at two numbers and to name the number that is larger. Lastly, the MN measure requires the participant to identify the missing number in a 3-consecutive number string within the 1-20 range (Clarke & Shinn, 2004).

These assessments are considered curriculum-based because they are aligned with the grade level curriculum with emphasis on expected computational skills. The system provides 40 alternative forms for each grade level to be used for Benchmark Assessment, Strategic Monitoring, and frequent Progress Monitoring. These components make it possible to test the interventions being used in a time-efficient way that is sensitive to change in performance (Shinn, 2004).

Generalizability, Validity, and Reliability of CBM

According to Chard et al. (2005), over the past 25 years hundreds of studies designed to screen students at-risk of reading difficulties have been conducted and published as compared to 50 or less mathematics studies designed for mathematics difficulties. The mathematics studies that have been done indicate that various EM-CBM assessments demonstrate reliability, validity, and sensitivity at the pre-school through first grade level (Baglici, 2008; Chard et al, 2005; Clarke & Shinn, 2004; Gersten & Jordan, 2005). Long-term studies that analyzed predictive validity including Jordan et al., (2009) and Jordan et al., (2010) found that pre-school through kindergarten EM-CBM are strongly correlated; evidenced by significant growth and accuracy of pre-school measures for identifying children in need of mathematics intervention (Baglici, 2008).

The earliest studies on early mathematics screening for mathematics difficulties began in the late 1980s. These studies mainly assessed children at a single point-in-time. In recent years, emphasis has been placed on the need for longitudinal studies of growth trajectories that provide a better understanding of mathematics difficulties and interventions at the preschool and kindergarten level (Gersten et al., 2005). The first longitudinal mathematics difficulties studies investigated the relationship between reading and mathematics difficulties. These researchers analyzed the nature of the deficits for different types of mathematics difficulties (Gersten et al., 2005). Contemporary tests that have been developed for preschool and kindergarten students include the Number Knowledge Test (NKT), Stanford Early School Achievement Test (SESAT), Number Sense Test (NST), Test of Early Mathematics Ability (TEMA), Preschool Math Curriculum-Based Measurement (PM-CBM), Kindergarten Number Sense Battery (KNSB), and Test of Early Numeracy (TEN).

EM-CBM studies use *number sense theory* to assess early mathematical skills. The assessments measure each student's ability to count, identify numbers, make judgments in quantity comparisons, and his/her use of a concrete and abstract number line (Baglici, 2008). For example, Clark and Shinn (2004) investigated the reliability, validity, and sensitivity of the early mathematics measures with first graders using the AimsWeb TEN. This study became a springboard for a multitude of others that replicated and expanded their work using other kindergarten and first grade mathematics tests and frequent data collection to measure student growth patterns for specific concepts and skills.

Concurrent validity. Clarke and Shinn (2004) assessed both concurrent and predictive validity for the EM-CBM TEN measures of OC, MN, QD, and NI. Correlations between those four EM experimental measures were examined against the outcome on three criterion assessments: Woodcock Johnson – Applied Problems (WJ-AP), NKT, and the M-CBM.

The intercorrelations among the experimental measures were high. OC had the lowest consistency (ranging from .55 to .79 with a median of .69) and therefore was excluded when measuring the other three measures. The range of NI measures in all three data collection periods was .72 to .93 with a median of .85. The range of QD measures was .86 to .93 with a median of .88 and the MN measures ranged from .72 to .90 with a median of .86.

Clarke and Shinn (2004) reported findings for the relations among the experimental and criterion measures in which QD measures had the strongest concurrent validity correlations ranging from .71 to .88 with a median of .75. The OC measures had the lowest concurrent validity correlations ranging from .49 to .70 with a median of .60. The NI measures ranged between .60 and .70 with a median of .66. The MN measure ranged from .68 to .75 with a median of .71.

To further investigate the concurrent validity correlations, Clarke and Shinn (2004) conducted a test of differences between dependent correlation coefficients. Their findings included the following:

- When comparing OC and QD, QD had significantly higher correlations with data collection of M-CBM scores in the winter assessment with t (49) = 2.93, $p < .05$ and spring with t (49) = 3.31, $p < .05$.
- When comparing NI and QD, QD had significantly higher correlations with data collection of NKT in the fall, t *(49) = 3.18,* p < .05.
- When comparing NI and QD, QD had significantly higher correlations with the spring data collection of M-CBM, t (49 = 3.03, $p < .05$.
- When comparing OC and MN assessment from the winter data collection, MN demonstrated a stronger relationship with M-CBM, t (49) = 3.50, $p < .05$.

Predictive validity. Predictive validity was assessed with the following measures (Clarke & Shinn, 2004):

1. Fall data collection of EM-CBM and winter data collection of M-CBM,

2. Spring data collection of WJ-AP and spring data collection of M-CBM,
3. Winter data collection of EM-CBM and winter data collection of WJ-AP and spring data collection of M-CBM.

All four of the experimental EM measures demonstrated strong relationships. The highest median correlation was a .76 for the QD measure. The median for MN was .76, NI was .68 and OC had .56 (Clarke & Shinn, 2004).

According to Clarke and Shinn (2004) other evidence of predictive validity was supported by comparisons between the strength of correlations for the EM-CBM by testing the differences between two dependent correlation coefficients.

1. In comparing QD and OC, QD had significantly higher correlation coefficients with M-CBM between fall and winter, t (49) = 3.18, $p < .05$ and between winter and spring, t (49) = 3.34, $p < .05$.
2. In comparing MN and OC, MN had significantly higher correlation coefficients with M-CMB between fall and winter, t (49) = 3.08, $p < .05$ and between winter and spring, t (49) = 3.34, $p < .05$.

Reliability. Using Pearson product moment correlation coefficients, Clarke and Shinn (2004) analyzed reliability of inter-scorer, alternate-form, and test and retest reliability of the early mathematics (EM) measures. The results were based on criteria on reliability analysis from Salvia and Ysseldyke (as cited in Clarke & Shinn, 2004):

a.) .90 or greater is recommended for making educational decisions for individual students;
b.) .80 or greater is recommended for making screening decisions for individual students;
c.) .60 or greater is recommended for making educational decisions for groups of students.

Clarke and Shinn (2004) used .80 as the reliability standard for their study based on the fact that this was a study for early identification of individual students without the involvement of high stakes tests and the fact that it did not change a student's grade level placement or educational classification.

Clarke and Shinn (2004) calculated the inter-scorer reliability by dividing the number of items that the two scorers agreed upon by the number of items they disagreed upon. This was done with the fall collection using 12 student protocols. The OC, NI, and QD measured at .99 reliability while the MN measured at .98; all surpassed the standard for making educational decisions.

The alternate form reliability was analyzed during the fall and winter sessions using three different alternate tests:

1. M-CBM grade 1 computation probes involving addition and subtraction,
2. WJ-AP subtest where students solved addition and subtraction problems,
3. NKT, which required students to work through levels of problems. The first level consisted of counting chips and identifying geometric shapes. The second level involved identifying bigger or smaller numbers and solving simple addition and subtraction problems. The third level required students to solve similar problems to level 2; however, the problems were made more difficult by using larger numbers. In addition, this level has problems that require students to state how many numbers are between a given pair.

The OC assessment for the EM-CBM measures was used for the alternate measures as well because the children were required to count orally for one minute in each assessment. Since the assessment was the same in each, the student was not asked to perform the test again. The

NI, QD, and MN alternative tests were done in different orders for students to avoid the practice effect (Clarke & Shinn, 2004).

The results of the reliability investigation found the OC, NI and QD all above the .90 standard for individual educational decision making. The MN was .83 in the fall and .78 in the winter so when averaged out it still exceeded the .80 standard for educational screening decisions on individual students (Clarke & Shinn, 2004).

The long-term test-retest reliability was examined with 13 weeks between fall and winter assessments and again with a 26-week period between fall and spring. Each of the EM-CBM measures came out above an acceptable .80 standard that is the minimum recommended for screening decisions made for individual students (Clarke & Shinn, 2004).

The study with Chard et al., (2005) extended Clarke and Shinn (2004) by administering TEN measures to 436 kindergarten students and 483 first grade students in the Pacific Northwest. This study included analyzing the early mathematics measures' sensitivity to growth by examining gains made from fall to spring. Although all four TEN tasks (OC, QD, MN, and NI) were used—modifications were made for kindergarten in that OC was only assessed in the fall, and the other three used numbers ranging from 1 to 20 in the fall but changed to 1 to 10 in the winter and spring. For the criterion assessment in the spring, the NKT was used based on prior research citing it as highly correlated with published measures of mathematics achievement in both kindergarten and first grade samples.

The Chard et al. (2005) replicated predictive and concurrent validity findings from the Clarke and Shinn (2004) study with kindergarten students. Patterns observed from the student growth data indicated that performance in the spring of kindergarten was higher on each task as compared to the fall of first grade. The same growth changes occurred in both kindergarten and first grade students when measuring QD, MN, and NI in the fall to winter to spring time spans; however, only NI task showed considerable change from fall to spring when compared to QD and MN.

Lembke et al. (2008) extended the previous research of Chard et al. (2005) with an emphasis on sensitivity to student progress over time. To monitor student growth, taken monthly rather than just two to three times a year as other studies had done. The results indicated that NI, QD, and MN had satisfactory alternate form reliability, with QD and NI having the stongest reliability coefficients with a range of .79 to .93 for kindergarten students. The strongest coefficients for the first grade students was observed in the QD task which ranged from .70 to .85. Concurrent validity was found by correlating the scores on the EM-CBM grade 1 standard scores of the SESAT and teacher ratings of mathematics performance. This data indicated that the validity were low to moderate ranging from .19 to .46 across the measures (Lembke et al., 2008).

Lembke et al. (2008) used a two-level hierarchical linear growth model to measure student progress over time while administering the MN, NI, and QD on a monthly basis. The results indicated that both kindergarten and first grade students showed a significant linear growth in NI by estimated weekly growth rates of .34 in kindergarten students and .24 in first grade students. The results for QD and MN were curvilinear which Lembke et al. (2008) indicated may suggest that either these measures are not good indicators of progress or students learn in “bursts of performance” as they progress in their learning. The strong results of the NI measure indicates that it may be a reliable tool for differentiating students within a specific grade level.

Lembke and Foegen (2009) conducted a study with 300 kindergarten and first grade students using QD, MN, NI, and quantity array to examine the reliability and validity of these single skill assessments. Lembke and Foegen (2009) administered the TEN tests to individual students as one-minute tasks, similar to Clark and Shinn (2004). The reliability results indicated the strongest coefficients for the NI, QD, and MN tasks with most correlations in the mid to high .80s. Validity findings were parallel with moderate to strong concurrent and predictive validity coefficients for the NI, QD, and MN tasks. The lowest correlations for both kindergarten and first grade were produced by the quantity array measure. This was a one-year study where students were tested in the fall, winter, and spring of the same school year.

Seethaler and Fuchs (2010) studied approximately 200 kindergarten and first grade students by creating and administering a group Computational Fluency assessment, an individually administered Number Sense test, and individually administered a single skill test for Quantity Discrimination. The purpose of the study was to examine the reliability, validity, and predictive use of EM-CBM assessments to screen students in kindergarten and first grade for indicators that would predict which students would be at-risk for mathematics difficulties.

Seethaler and Fuchs (2010) also compared the results between the multi-screen assessments and the single-skill assessment for QD. Reliability findings showed coefficients for the average of the two multi-skill screeners were .90 for both the fall and spring administrations. Concurrent and predictive validity for the multi-screeners in respect to the fall kindergarten to spring first grade measures ranged from .55 to .72, which was closely aligned with the QD that varied from .52 to .66. In addition, there were no significant differences in predictive utility when comparing the single-skill (QD) and multi-skill tests that take much longer to administer.

However, Seethaler and Fuchs noted that the multi-skill assessments provided more information to teachers than the single-skill assessment of just QD.

Jordan et al. (2009) studied the predictive relationship between early number competence in kindergarten and later mathematics achievement in third grade student performance on criterion assessments. This study used the Number Sense Brief (NSB) to measure 204 kindergarteners' number sense and measured the same children again in the beginning of first grade. NSB is a research-based untimed multiple-skills assessment that takes approximately15 minutes to administer and consists of 33 number sense items.

Based on the results of the NSB, strong predictability was shown to correlate to performance on the Woodcock Johnson Calculation and Applied Problems (WJ-AP) assessment at the end of both first grade and third grade. This also was the outcome of the correlation when comparing a fourth year of data on the same students generated from the Delaware Student Testing Program (DSTP), the state criterion high stakes test. Repeated measures analysis of variance (the test was conducted 7 times throughout the year) showed a statistically significant main effect for the group which revealed that children who met the DSTP mathematics standard at the end of third grade consistently obtained higher NSB scores across time than those who did not meet the mathematics standard.

The kindergarten early number competency battery consisted of 42 items that included tapped counting and number recognition, number comparisons, nonverbal calculation, story problems, and number combinations. The WJ-AP was used to measure third grade composite mathematics scores. According to Jordan et al. (2009), the WJ-AP demonstrated high content validity and was significantly correlated with performance on the grade 3 DSTP.

Over the four-year period the students were assessed 11 times between kindergarten and grade 3. The number measure was administered four times in kindergarten and two times in the beginning of their first grade year. The mathematics achievement assessments were conducted five times—the spring of first grade and the fall and spring of both second and third grade. The results of the study indicated that all of the correlations between the number competence measures and the mathematics achievement scores were positive and significant ($p<.01$).

As was noted, mathematical EM-CBM is in its infancy compared to the work completed in reading. More mathematical research needs to be done to pinpoint early numeracy skills that most critically build a foundation for later hierarchical mathematics studies and reinforce current research that supports identifying students who would benefit from intervention (Baglici, 2008; Clarke & Shinn, 2004; Jordan, Glutting, & Ramineni, 2009).

The current study was designed to expand the one year kindergarten study of Clarke and Shinn (2004) to a 4 year kindergarten to third grade study to investigate concurrent and predictive validity among the individual TEN measures, M-CBM assessments, and the grade 3 MontCAS.

Chapter 3

Methodology

This chapter begins with a demographic overview of the participating school district. In addition, a description of the participants, measures, research design, data collection and data selection processes are explained. A presentation of the data analysis procedures utilized concludes this section.

Population and Participant Selection Process

The school district involved in this study was located in a rural area of Montana with agriculture, timber, and tourism central to the economy. The district serves approximately 264 students in kindergarten through eighth grade and approximately 129 high school students. Enrollment has declined by 73 students over the past 5 years. Ninety percent of the students are Caucasian with 10% being of other races (not specified).

In the overall district, 36% of the student population qualified for free and reduced lunch. District percentage of IDEA students has dropped slightly over the last four years and holds at 12% while the graduation rate has been at 98% over the past 4 years.

This district was selected because of the number of years that the AIMSweb has been part of the grade K-6 curriculum. During the 2005-2006 school year, the district began utilizing AIMSweb assessment system and continued to do so throughout the time of this study.

The district adopted the McGraw Hill textbook series for the 2007-2008 school year but the textbook was not a major part of the primary grade curriculum during this study. The core mathematics program includes doing Calendar Math and Rocket Math each day in all classrooms. The district began developing their own mathematics curriculum during the last year of the study with an instructional emphasis on concrete-representational-abstract (C-R-A) methods for teaching all math concepts at all grades. The focus for mathematics content is driven by the Common Core Standards, the NCTM Focal Points, and the MontCAS. The district has combined these targets to address their students' specific needs.

The district's most widely used math intervention is Number Worlds. In addition, supplements include Corrective Math (multiplication & division), Plato (computer-based program), and Skill Builders. The main intervention work in mathematics during this study focused on re/pre teaching using the C-R-A method for students identified as needing intervention. The district also began implementing a 90 minute mathematics block in the daily schedule during the last two years of the study.

Process Used to Collect Data

The data were retrospectively attained through the district in accordance with The University of Montana Institutional Review Board and the standards required by the participating school district. Permission was attained by the district elementary principal who provided an anonymous database with students' names replaced with a unique identification number. This anonymous data bank did not identify any other demographics including gender, social economic background, or age. The only data provided included a unique identification number, available TEN fall, winter, and spring assessment scores for kindergarten and first grade students along with available first, second, and third grade M-CBM fall, winter, and spring scores, and available grade 3 MontCAS scores.

For statistical purposes, two cohorts of students were combined by using the kindergarten group from the 2005-2006 school year and comparing their scores with the kindergarten 2006-2007 group during their kindergarten through grade 3 years. The available data showed 12 students with scores in the 2005-2006 kindergarten class and 31 in the 2006-2007 kindergarten class for a total of 43 students. The data were screened based on the criteria that participant information needed to include grade 3 MontCAS scores. Eight students from kindergarten 2005-2006 fit the criteria as did 21 students from the 2006-2007 kindergarten class. Twenty-nine students who had data from kindergarten to grade 2 and a grade 3 MontCAS score were included in the study. Although some students were missing one or more kindergarten and/or first grade AIMSweb assessments but had a recorded MontCAS score they were considered to have met the criteria to be included in the study.

An independent-sample t-test was conducted to compare the available TEN and M-CBM scores for the non-eligible students with the participating students. The following table compares the descriptive statistics and t-scores that were calculated.

Table 1

Descriptive Statistics Comparing K-2 AIMSweb Scores of non-Eligible and Participating Groups

	Non-Eligible		**Participating**		**Difference**		**t-score**
Assessment	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Gr2 CBM	17	7	21	7	4	0	1.2
K MN	11	5	14	5	3	0	1.6
Gr1 CBM	10	4	13	4	3	0	1.7
Gr1 MN	17	3	20	5	3	2	1.7
Gr1 QD	32	10	38	8	6	2	1.8
K QD	18	8	24	7	6	1	1.9
Gr1 NI	59	18	71	12	12	6	2.1
K NI	39	13	50	12	11	1	2.1
Gr1 OC	77	9	91	13	14	4	2.7
K OC	41	17	64	18	23	1	3.1

Note. Gr2 = grade 2, K = kindergarten, Gr1 = grade 1, M-CBM = Mathematics Curriculum Based Measures, MN = Missing Number, QD = Quantity Discrimination, NI = Number Identification, OC = Oral Counting.

In every assessment comparison, the non-participating group had a lower average score than the participants. Figure 1 illustrates the outcome of the t-score procedure conducted for the comparison of the two groups.

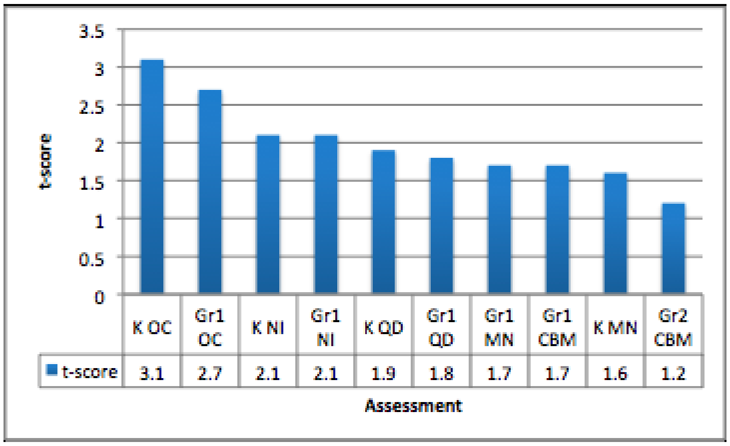


Figure 1. Non-eligible and Participating Students' t-scores on AIMSweb Assessments Scores. Gr2 = grade 2, K = kindergarten, Gr1 = grade 1, M-CBM = Mathematics Curriculum Based Measures, MN = Missing Number, QD = Quantity Discrimination, NI = Number Identification, OC = Oral Counting.

The assessments with the least similarity in scores were the Oral Counting and Number Identification for both kindergarten and first grade. For example, the kindergarten Oral Counting assessment showed that the non-eligible students scored on the average 23 points less than the participating students. The t-score of 3.1 informs the research that very few scores in the non-eligible group would be found in the set of participating students' data. This is to say that the scores of those that were not eligible were substantially less on the average than the scores of the students who were found to be eligible participants in Oral Counting and Number Identification.

On the other hand, the non-eligible kindergarten students in the Missing Number subtest scored 3 points less on the average than the participating group. The t-value of 1.7 suggests that scores between the non- eligible group and the participating group co-mingle to a greater extent than in subtests that have larger t-scores. This can be said about each of the subtests except the Oral Counting and Number Identification as stated above.

The next step of data selection involved a comparison of the participating students' scores from the two cohorts who included the 2005-2006 kindergartners through their third grade year (first cohort) and the 2006-2007 kindergarteners through their third grade year (second cohort). It was found that the two groups performed at comparable levels and were combined in order to have a larger population. There were eight students from the kindergarten 2005-2006 cohort and 21 from the 2006-2007 cohort that had grade 3 MontCAS scores; however, some of those students were missing one or two scores from the fall, winter, and/or spring AimsWEB assessments. These students were still included based on the fact that the study included 18 AIMSweb assessments and so the available scores provided valuable information to the results of the study.

In order to set up the comparison, the mean scores of each assessment were calculated by averaging the fall, winter, and spring TEN scores, which included Oral Counting (OC), Number Identification (NI), Quantity Discrimination (QD), and Missing Number (MN) for grades kindergarten and grade one. A mean was also calculated for the fall, winter, and spring scores for the Mathematics-Curriculum Based Measures (M-CBM) conducted in grades 1 and 2 in order to correlate them with the grade 3 Montana Comprehensive Assessment System (MontCAS).

Figure 2 demonstrates that the two cohorts were comparable through similar performance on each assessment. The MontCAS is scored using a scaled system where raw scores are grouped as indicated in Table 2 (Montana Office of Public Instruction [OPI], 2011). The analysis of the data utilized the scaled scores that were provided by the school district. Raw scores were not available.

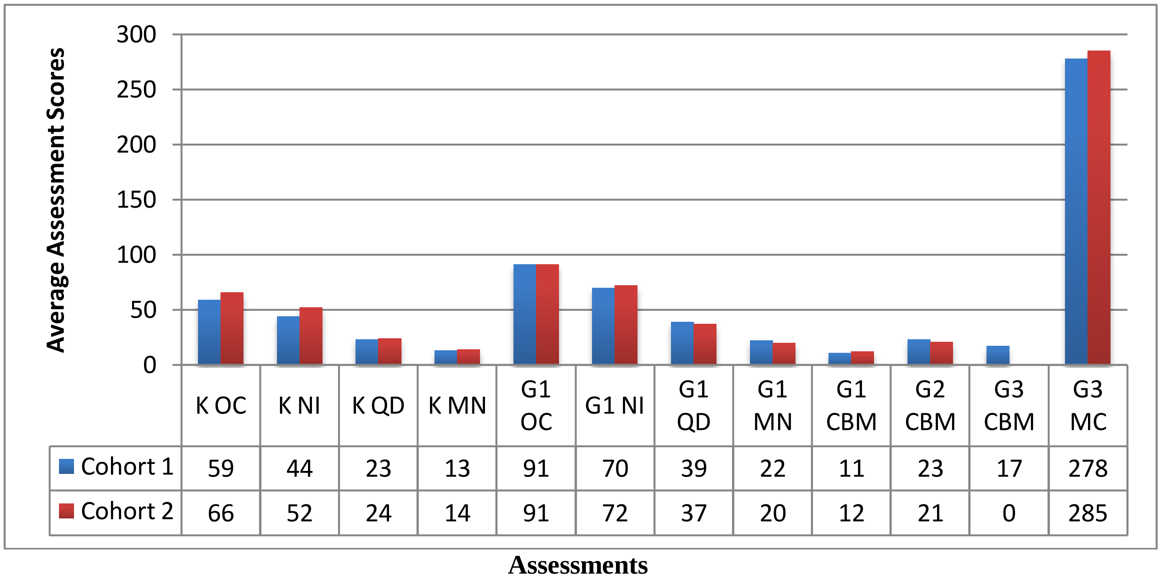


Figure 2. Comparison of Tests of Early Numeracy (TEN) for Cohorts 1 and 2. K = kindergarten, G1 = grade 1, OC = Oral Counting, NI = Number Identification, QD = Quantity Discrimination, and MN = Missing Number, CBM = Math Curriculum Based Measures, and G3 MC = Grade 3 Montana Comprehensive Assessment System. Cohort 1: N = 8 and Cohort 2: N = 21.

Each TEN and M-CBM average includes the fall, winter, and spring scores. In some cases, the student may have missed one assessment score for fall, winter, or spring. The second cohort was not given the grade 3 M-CBM because the District had upgraded the AIMSweb system in grade 1 through 8 assessments to Math Concepts and Applications (M-CAP). The M-CBM assessed computational fluency and the M-CAP assessed problem solving so only the MontCAS score was used in the study for grade 3.

The comparison of the groups demonstrated that it was reasonable to combine the two cohorts to investigate the questions for this study. The mean of the fall, winter, and spring assessments indicated the comparable levels of performance throughout the four years of available data that was analyzed.

Instruments Used to Collect Data

Kindergarten and early first grade participants' mathematics skills were assessed with the TEN. These measures were the independent variables and experimental measures that consisted of four number sense measures that include Oral Counting (OC), Number Identification (NI), Quantity Discrimination (QD), and Missing Number (MN).

The OC measure required the participant to begin with 1 and orally count for one minute. The participant used no classroom materials. The recorder held a sheet with numbers and marked any numbers that were skipped. If a participant struggled for 3 seconds, the recorder said the next number. The participant's score indicated the number of correct numbers counted in that one minute and ranged from 0 to over 100, depending on how many numbers the student counted in that single minute.

The NI measure provided an 8x7 grid of randomly listed numbers ranging from 1 to 20. The participant was required to orally name the numbers from left to right in the rows. If the participant hesitated for 3 seconds, he/she was told to go to the next number. After one minute, a score was calculated by counting the number of correct responses during that minute of time. Fifty-six (56) test items appeared on this measure; therefore, the range of scores possible was 0 to 56.

The QD measure required the participant to be provided with a grid of 28 individual boxes with two different numbers in each box ranging from 1 to 10 for kindergarteners and 1 to 20 for first graders. The participant was asked to begin with the left box in the top row and name the number in each box that is larger. The participant's score was based on how many boxes he/she correctly identified and named the higher number. When a participant hesitated for

3 seconds, he/she was told to try the next box. The assessment contained 28 items; therefore, the possible range of scores is from 0 to 28.

The MN measure required the participant to be provided with a grid of 21 boxes each of which had a string of 3 consecutive numbers; however, one number was missing. The numbers ranged from 0 to 20 and the participants were given one minute to identify the missing number in each beginning at the top and proceeding. A score of one point was given for each correct response. If a participant hesitated for 3 seconds, he/she was told to try the next box. There were 21 items on this assessment; therefore, possible scores ranged from 0 to 21.

The formative assessment used for grade 1 through grade 6 was the AIMSweb M-CBM and Math Fact Probes. These assessments are considered curriculum-based because they are aligned with the relative grade level curriculum. The program's emphasis on computation is based on expected computational skills for each grade level. The system provides 40 alternative forms for each grade level use for Benchmark Assessment, Strategic Monitoring, and frequent Progress Monitoring. These components make it possible to test the interventions being used in a time efficient way that is sensitive to improvement (Shinn, 2004).

The assessments were made up of a wide-range of computation problems aligned with grade level curriculum. The problems are narrow-band tests meaning that they consist of many problems. Students typically had two to four minutes to complete the test depending upon grade level. Common to the EM-CBM, the assessment was based on what the student did correctly. The score was derived by the number of Correct Digits written instead of the correct answer as a whole (Shinn, 2004). These assessments were conducted each fall, winter, and spring for the district and those available scores were used for this study.

The Montana Comprehensive Assessment System (MontCAS) includes the Criterion

Referenced Test (CRT) that was administered in grades 3 through 8 and grade 10 each spring in Montana accredited schools. The mathematics CRT consisted of word problems that were meant to demonstrate the students' thinking; therefore, students were required to write computation and in some instances devised tables and charts in the responses. The CRT was intended to be a measure of student proficiency rather than speed; therefore, suggested times were given but students were allowed to continue as long as they were working productively. The suggested time for grades 3-8 mathematics CRT was 45 to 55 minutes (Montana Office of Public Instruction, 2010).

The Montana Comprehensive Assessment System (MontCAS) is a criterion test that is administered to students in grade 3 through 8 and grade 10 across the state. As can be seen in Table 2, the raw scores are scaled and divided into categorical levels with *Proficient* being the minimal benchmark set for defining grade level performance.

Table 2

Mathematics MontCAS Raw to Scaled Score Ranges Defining Performance Level

Raw Score Ranges			
2008-2009	2009-2010	Scaled Scores	Performance Level
54-66	53-66	290-300	Advanced
41-53	41-52	250-289	Proficient
33-40	33-40	225-249	Nearing Proficiency
0-32	0 - 32	200-224	Novice

Note. MontCAS = Montana Comprehensive Assessment System.

The scaled scores were those that determined the level of performance each student had attained and the scores that were available for the study. The raw scores are unevenly distributed with nearly half included in the Novice level.

Reliability and Validity of Instruments

The assessments were conducted by trained staff in each classroom. Data entry was conducted by the elementary principal who was also trained in AIMSweb. The data were analyzed to investigate correlations for each of the EM-CBM's Test of Early Numeracy (TEN) measures (OC, NI, QD, and MN) in kindergarten and grade 1 with the M-CBM in grades 1 and 2 along with the CRT grade 3 MontCAS scores. The study looked for predictive validity and significance of the relationship between the participants' performance in K-1 early measures and the same participants' performance on grade 3 MontCAS scores.

Research Design

This quantitative study implemented a longitudinal design to measure student performance over a four-year period. The study measured student performance from kindergarten (year 1) through grade 3 (year 4). As was stated earlier, two cohorts were used that included the kindergarten class of 2005-2006 and the 2006-2007 kindergarten class. The data collected included three administrations of kindergarten (year 1 and year 2) TEN measures. This took place during the fall, winter, and spring of the year with approximately 13 weeks in-between. During these same students' first grade year, the TEN measures were administered again during the fall, winter, and spring of the school year approximately 13 weeks apart. In addition, the M-CBM was administered three times throughout the year (fall, winter, and spring) to each cohort during grades 1 and 2. Finally, when participants were in grade 3, the MontCAS was administered in the spring.

Procedures

Prior to beginning data collection, permission was granted from the district and The University of Montana Institutional Review Board (IRB) for conducting research. Immediately after acquiring permissions from the IRB and district office, authorization from the elementary principal/test coordinator was granted and the retrospective data were provided through a spreadsheet with student names omitted and substituted with unique identification numbers.

Analysis Procedures

Data analysis procedures used in this study included descriptive statistics, Pearson product moment correlations, and sequential regression. Descriptive statistics were used to find means and standard deviations that describe participant characteristics and overall performance on each of the measures.

Correlations and regression equations were utilized to answer the research questions regarding the predictive validity of kindergarten TEN performance for first grade TEN and M-CBM performance. This was done by calculating Pearson product moment correlations for all of the data as an initial step. Regression analysis was used to investigate a relationship between kindergarten TEN performance, first grade TEN performance, grades 1 and 2 M-CBM and grade 3 MontCAS scores. The kindergarten and grade 1 TEN scores, along with grade 1 and 2 M-CBM scores were used as predictor variables and MontCAS scores were used as the outcome variables.

Delimitations. Students who had not attended the district from kindergarten through the first grade and/or did not have recorded grade 3 MontCAS scores were not included in the study. For this reason, the study represents a more stable population than the district overall.

Limitations. Fidelity of testing is one limitation. The researcher is using data submitted by the classroom teachers who implemented the assessments. The researcher is making the assumption that all testing regulations were followed consistently by the different individuals conducting the tests from year to year.

Summary

The methodology design for this research consisted of a quantitative design that allowed the researcher to investigate the correlation of each TEN measure in kindergarten and grade 1 and M-CBM in grades 1 and 2 with the criterion referenced MontCAS tests in grade 3. The study explored the predictive validity as well as the significance of the relationship between the participants' performance in kindergarten and first grade on each of the TEN and grades 1 and 2 M-CBM tasks with the grade 3 MontCAS scores.

Participants for this study included 29 kindergarten through grade 3 students from a school district located in rural Montana. Only students who had at least one assessment score from each grade level were included in the 4-year longitudinal study.

Chapter 4

Results

Introduction

A minimal number of studies in early mathematics interventions have taken place over the past 25 years when compared to the hundreds of studies completed for screening students at-risk of reading difficulties (Chard et al., 2005). Modeled after studies done in reading, more and more long-term mathematics studies for preschool through grade 1 have been conducted and have shown strong correlations to statistically significant growth and accuracy in student mathematics performance (Gersten et al., 2005). These studies play a crucial role in developing scientifically based assessments that will allow educators to identify students who are at risk of struggling in mathematics at the preschool through grade 1 level (Jordan et al., 2009; Jordan et al., 2010).

For these reasons, this study was designed to investigate whether there is a longitudinal association between performance on kindergarten and first grade Tests of Early Numeracy (TEN), grade 1 and grade 2 Mathematics Curriculum Based Measurement (M-CBM) and the same students' performance on the grade 3 Montana Comprehensive Assessment System (MontCAS).

This chapter reports and summarizes the findings based on the correlation and regression statistics explored through the experimental (TEN and M-CBM) and the criterion (MontCAS) variables. Descriptive statistics include the mean and standard deviation for the fall, winter, and spring testing periods for the experimental measures and the grade 3 testing for the criterion measure. Second, the predictive validity of each TEN measure is illustrated with charts and then discussed. Appendix B provides scatter plots to illustrate the linear regression for each of the

AIMSweb assessments with the grade 3 MontCAS. Third, a correlation matrix is presented to examine the relationship among the TEN and M-CBM assessments and MontCAS scores. Fourth, sequential multiple regression analysis was used to assess the relationship among the kindergarten and grade 1 TEN measures built upon with grades 1 and 2 M-CBM, and the MontCAS assessment.

Data

The data used in this study includes two cohorts of kindergarten through grade 3 students from a single rural school district. The population that met specific criteria (a grade 3 MontCAS score) was used from the two cohorts who included the 2005-2006 kindergartners through their third grade year (first cohort) and the 2006-2007 kindergarteners through their third grade year (second cohort). The two classes were combined for an N of 43 cases; however, 14 students did not meet the eligibility criteria because they were missing grade 3 MontCAS scores and were screened out. The elimination involved four of 12 students from the first cohort and 10 of 31 students from the second cohort leaving 29 cases to be included in the study.

The eliminated students' available kindergarten and grade 1 TEN scores along with available grade 1 and 2 M-CBM scores were compared to those of the participating group. The eliminated students scored below in each subtest especially in kindergarten and grade 1 Oral Counting (OC) and Missing Number (MN). The rest of the assessment scores between the two groups co-mingled to a greater extent (see Table 1).

To demonstrate that it was reasonable to combine the two cohorts of participating students, the mean scores of each assessment were calculated by averaging the available fall, winter, and spring TEN scores which included OC, NI, QD, and MN for grades kindergarten and grade one. A mean was also calculated for the available fall, winter, and spring scores for the M-

CBM conducted in grades 1 and 2 in order to correlate them with the grade 3 MontCAS. When a test score from fall, winter, or spring was not available, no mean score was recorded and the participant was not included in the overall average for that assessment. Figure 2 demonstrates that the two cohorts were comparable through similar performance on each assessment based on the fact that some of the 29 students were not included as indicated in the Descriptive Analysis charts below and explanations that follow each chart.

Descriptive Analysis

First the data were sorted and organized by cohort based on the year each group began kindergarten. As was stated above, the cohorts that fit the study included the 2005-2006 class and 2006-2007 kindergarten classes. The researcher then examined the data for patterns and trends which included sorting the participants into performance levels on the MontCAS. The following figure illustrates the level of performance of the participating students.

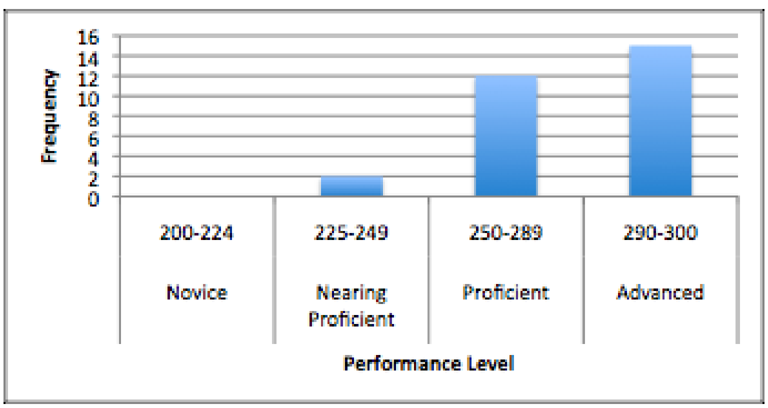


Figure 3. Grade 3 performance on the Montana Comprehensive Assessment System (MontCAS).

The shape of the histogram above shows that most scores are found at the Advanced and Proficient performance levels respectively with no scores at the Novice level. The fairly homogeneous distribution is skewed to the left. The results of these findings indicate that the group is the not- at-risk students and will be referred to as such from here on in the analysis. Table 3 provides descriptive statistics for the scores.

Table 3

MontCAS Scores Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
MontCAS	29	226	300	283.3	21.5

Note. MontCAS = grade 3 Montana Comprehensive Assessment System

A longitudinal analysis was done for each of the TEN and M-CBM formative assessments that were scored during the fall, winter, and spring of grades kindergarten through 3. The TEN assessments were utilized for kindergarteners and first graders. The M-CBM was used in grades 1 through 6 up until the first cohort completed grade 3 and then a different assessment was used for grades 1 through 6. For this reason, the current study was able to use only grades 1 and 2 M-CBM scores. The following tables and charts illustrate the mean, range, and standard deviation for each implementation of the assessments.

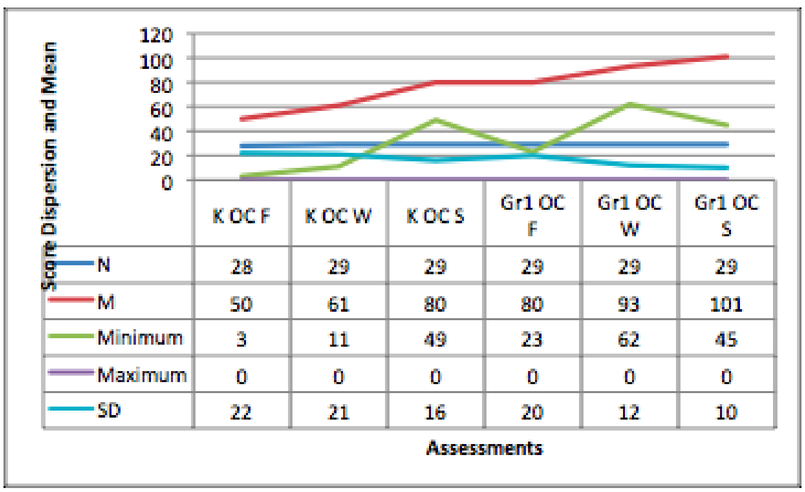
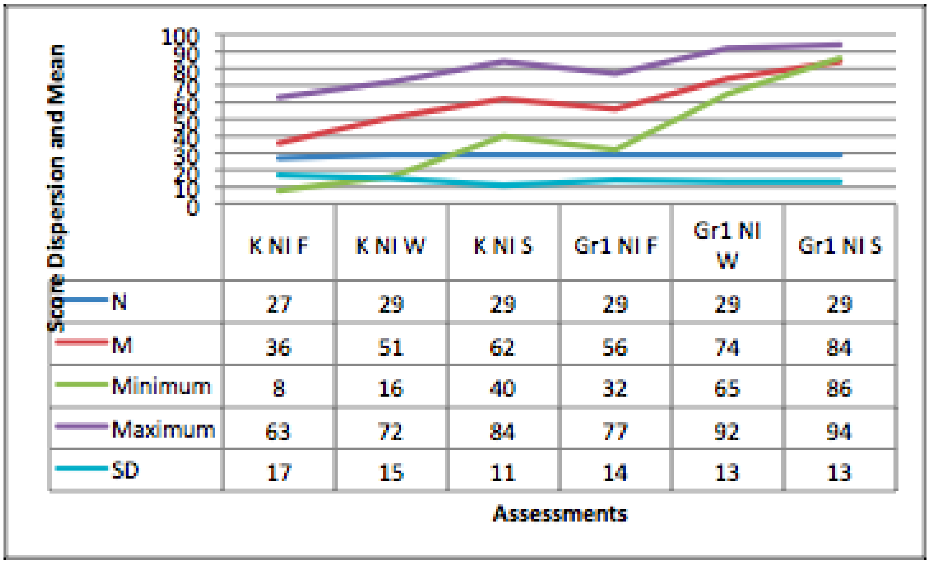


Figure 4. Longitudinal descriptive analysis on Oral Counting (OC). K = kindergarten, Gr1 = grade 1, F = fall, W = winter, S = spring, N = number of available scores, M=mean score, Minimum = minimum score, Maximum = maximum score, SD = standard deviation.

The results of the oral counting longitudinal analysis indicated continual growth for students between kindergarten and grade 1. The mean doubled from the kindergarten fall assessment to the grade 1 spring assessment with a plateau between the kindergarten spring and grade 1 fall scores. The standard deviation results indicated that the distribution of scores decreased throughout each year from fall to spring. The distribution of scores decreased by more than half between kindergarten fall and grade 1 spring.



	K NI F	K NI W	K NI S	Gr1 NI F	Gr1 NI W	Gr1 NI S
N	27	29	29	29	29	29
M	36	51	62	56	74	84
Minimum	8	16	40	32	65	86
Maximum	63	72	84	77	92	94
SD	17	15	11	14	13	13

Figure 5. Longitudinal descriptive analysis on Number Identification (NI). K = kindergarten, Gr1 = grade 1, F = fall, W = winter, S = spring, N = number of available scores, M = mean score, Minimum = minimum score, Maximum = maximum score, SD = standard deviation.

The results of the Number Identification longitudinal analysis indicated continual growth for students between kindergarten and grade 1. The mean more than doubled from the kindergarten fall assessment to the grade 1 spring assessment with a decrease between the kindergarten spring and grade 1 fall scores. The range and standard deviation results indicated that the largest distribution of scores was in the kindergarten fall assessment and smallest in the kindergarten spring assessment.

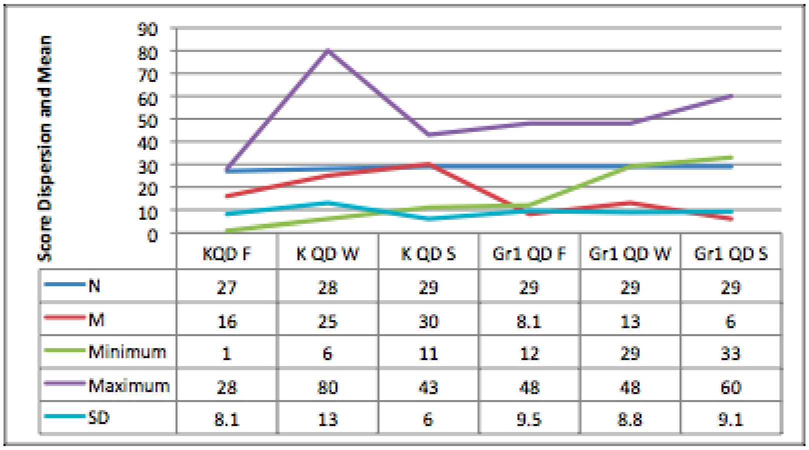
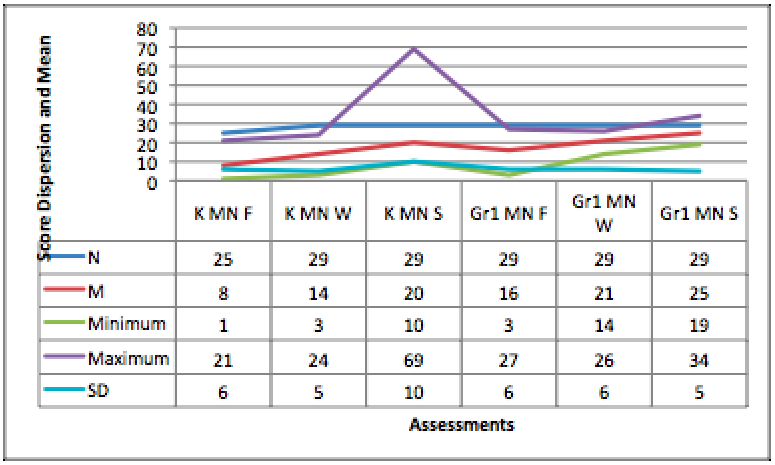


Figure 6. Longitudinal descriptive analysis on Quantity Discrimination (QD). K = kindergarten, Gr1 = grade 1, F = fall, W = winter, S = spring, N = number of available scores, M = mean of scores, Minimum = minimum score, Maximum = maximum score, SD = standard deviation.

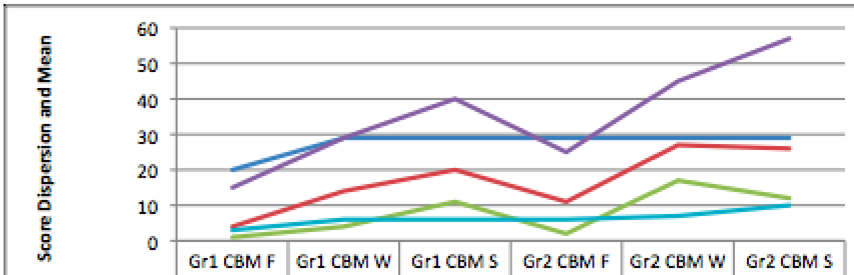
The results of the Quantity Discrimination longitudinal analysis indicated continual growth for students between kindergarten and grade 1. The maximum kindergarten score in the winter was over twice as high as the fall and nearly twice as high as the spring maximum. The minimum score rose steadily with a stronger gain in between grade 1 fall and winter minimum score. The mean nearly tripled from the kindergarten fall assessment to the grade 1 spring assessment with a plateau between the kindergarten spring and grade 1 fall scores. The standard deviation results indicated that the largest distribution of scores was in the kindergarten winter assessment which was also with one student score missing and smallest in the kindergarten spring scores when all students had available scores. The first grade distribution and N stayed the same with each assessment throughout the year.



	K MN F	K MN W	K MN S	Gr1 MN F	Gr1 MN W	Gr1 MN S
N	25	29	29	29	29	29
M	8	14	20	16	21	25
Minimum	1	3	10	3	14	19
Maximum	21	24	69	27	26	34
SD	6	5	10	6	6	5

Figure 7. Longitudinal descriptive analysis on Missing Number (MN). K = kindergarten, Gr1 = grade 1, F = fall, W = winter, S = spring.

The results of the Missing Number longitudinal analysis indicated continual growth for students between kindergarten and grade 1. The maximum kindergarten score in the spring was nearly three times higher than the winter maximum score. The minimum score rose steadily except for a drop between the kindergarten spring and grade 1 fall scores. The mean tripled from the kindergarten fall assessment to the grade 1 spring assessment with a decrease between the kindergarten spring and grade 1 fall scores. The standard deviation results indicated that the smallest distribution of scores was in the kindergarten winter and grade 1 spring assessments. Grade 1 scores showed a steady distribution level throughout the year with the smallest variation in the spring based on the range and standard deviation.



	Gr1 CBM F	Gr1 CBM W	Gr1 CBM S	Gr2 CBM F	Gr2 CBM W	Gr2 CBM S
N	20	29	29	29	29	29
M	4	14	20	11	27	26
Minimum	1	4	11	2	17	12
Maximum	15	29	40	25	45	57
SD	3	6	6	6	7	10

Assessments

Figure 8. Longitudinal descriptive analysis on Mathematics Curriculum Based Measurement (CBM). K = kindergarten, Gr1 = grade 1, F = fall, W = winter, S = spring, N = number of available scores, M = mean of the scores, Minimum = minimum score, Maximum = maximum score, and SD = standard deviation.

The results of the Mathematics-Curriculum Based Measurement (M-CBM) longitudinal analysis indicated continual growth for students during each school year; however, there were significant decreases between the kindergarten spring and grade 1 fall scores. The decreases are also evident in the minimum and maximum scores. In addition, each school year showed the mean multiplied from the fall assessments to spring assessments. The standard deviation results indicated that the smallest distribution of grade 1 scores was in the fall and stayed steady for the remainder of the year. The distribution of Grade 2 scores slightly rose from fall to spring.

Statistical Analysis

The first four research questions investigated the predictive validity of each kindergarten and first grade TEN experimental measure. Predictive validity of each measure was examined to see how well it predicted the grade 3 score on the MontCAS criterion referenced test. According to Cohen (1992) the criteria for effect sizes, small, medium and large coefficients are measured by an r of .10, .30, and .50 respectively. These criteria were used in analyzing the following correlations calculated with SPSS software. Variables for each case included the individual fall, winter, and spring scores along with a separate variable for the average of each student's three scores in each grade level. For students who were missing one of the three scores, the average score was not calculated as it would not have been accurate since noted on the tables below, the measured scores increased throughout each school year. For example, if a student was missing a fall score, the average score for fall, winter, and spring was also missing because an average of the winter and spring scores alone would not be accurate to be considered with the rest of the N's fall, winter, and spring score averages. In addition, when processing the analysis procedures with SPSS, the option to “exclude cases pairwise” was chosen in dealing with missing data. This option excluded the cases only if they were missing required data for the specific analysis being run. On the other hand, the same cases were included in any of the analyses that they had all required information (Pallant, 2007).

Questions 1: Does an Oral Counting test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?

To address research question 1, the following correlation table was prepared by including the individual scores for each administration of the measure followed by the average score for fall, winter, and spring as a separate variable.

Table 4

Correlation of Kindergarten and Grade 1 Oral Counting and Grade 3 MontCAS Scores

Grade Level and Test Interval	Pearson r	Pearson r^2
Kindergarten Oral Counting Fall	.44	20%
Kindergarten Oral Counting Winter	.38	14%
Kindergarten Oral Counting Spring	.40	16%
Kindergarten Oral Counting Average Fall, Winter, Spring	.44	20%
Grade 1 Oral Counting Fall	.16	2%
Grade 1 Oral Counting Winter	.04	0%
Grade 1 Oral Counting Spring	.09	1%
Grade 1 Oral Counting Average Fall, Winter, Spring	.09	1%

Note. One kindergarten student had a missing fall score.

Oral Counting predictive validity correlations ranged from $\underline{r}$ = .04 in the winter grade 1 scores to .44 in the fall kindergarten scores. Therefore, all kindergarten scores were medium predictors of Grade 3 MontCAS scores but decreased to small for all first grade scores. When calculating the correlation of the mean of the three kindergarten scores, the correlation stays at the highest level throughout the kindergarten year at .44. In first grade the correlation remained small for individual and calculated mean scores. Based on these findings, a mean of 44% of the variance of the kindergarten Oral Counting variable explained its linear relationship with Grade 3 MontCAS scores and a mean of 1% of the grade 1 Oral Counting variable explained its linear relationship with the Grade 3 MontCAS scores. Considering these data, the kindergarten scores would require one to reject the hypothesis (H_{01}) that there is no statistical relationship between

OC and the grade 3 MontCAS because it showed a moderate association; however, the grade 1 scores would require one to support the hypothesis.

Questions 2: Does a Number Identification test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?

To address research question 2, the following correlation table was prepared by including the individual scores for each administration of the measure followed by the average score for the year as a separate variable.

Table 5

Correlation of Kindergarten and Grade 1 Number Identification and Grade 3 MontCAS Scores

Grade Level and Test Interval	Pearson r	Pearson r^2
Kindergarten Number Identification Fall	.56	31%
Kindergarten Number Identification Winter	.58	33%
Kindergarten Number Identification Spring	.49	24%
Kindergarten Number Identification Fall, Winter, Spring Average	.62	39%
Grade 1 Number Identification Fall	.50	25%
Grade 1 Number Identification Winter	.39	15%
Grade 1 Number Identification Spring	.33	11%
Grade 1 Number Identification Fall, Winter, Spring Average	.47	22%

Note. Two kindergarten students had missing fall scores.

Number Identification predictive validity correlations ranged from r = .33 in the spring of grade 1 to .58 in the winter of kindergarten. The kindergarten year correlation for the average scores exceeded the individual test correlations and came out with a mean of 39% explained variance of the Number Identification variable's linear relationship with Grade 3 MontCAS

Scores. Grade 1 scores showed a large correlation in the fall but continually decreased to a medium correlation and indicated a 16% mean of grade 1 variance of the Number Identification variable with explained linear relationship of Grade 3 MontCAS scores. Considering these data would require one to reject the hypothesis (H_{02}) that there is no statistical relationship between NI and the grade 3 MontCAS for both kindergarten and grade 1.

Question 3: Does a Quantity Discrimination test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?

To address research question 3, Table 6 was prepared by including the individual scores for each administration of the measure followed by the average score for the year as a separate variable.

Table 6

Correlation of Kindergarten and Grade 1 Quantity Discrimination and Grade 3 MontCAS Scores

Grade Level and Test Interval	Pearson r	Pearson r^2
Kindergarten Quantity Discrimination Fall	.44	19%
Kindergarten Quantity Discrimination Winter	.48	25%
Kindergarten Quantity Discrimination Spring	.42	19%
Kindergarten Quantity Discrimination Fall, Winter, Spring Average	.59	35%
Grade 1 Quantity Discrimination Fall	.42	22%
Grade 1 Quantity Discrimination Winter	.19	4%
Grade 1 Quantity Discrimination Spring	.37	14%
Grade 1 Quantity Discrimination Fall, Winter, Spring Average	.39	15%

Note. Two kindergarten students had missing fall scores and one grade 1 student had a missing winter score.

Quantity Discrimination predictive validity correlations ranged from $\underline{r} = .19$ in the winter grade 1 scores to .48 in the winter kindergarten scores. All three kindergarten scores were medium predictors of Grade 3 MontCAS scores; however, when using the mean variable of those score' predictability rose to large at .59. Therefore, the mean kindergarten scores indicated 35% of the explained variance of this variable. First grade scores ranged from a small correlation of .19 for the winter administration of the assessment to a medium correlation for fall and spring. The mean grade 1 scores indicated 15% explained variance of Quantity Discrimination. Again these data require one to reject the hypothesis (H_{03}) that there is no statistical relationship between QD and the grade 3 MontCAS in grades kindergarten and 1.

Questions 4: Does a Missing Number test of early numeracy in kindergarten and grade 1 correlate with mathematics performance in grade 3?

To address research question 3, Table 7 was prepared by including the individual scores for each administration of the measure followed by the average score for the year as a separate variable.

Table 5

Correlation of Kindergarten and Grade 1 Missing Number and Grade 3 MontCAS Scores

Grade Level and Test Interval	Pearson r	Pearson r^2
Kindergarten Missing Number Fall	.52	27%
Kindergarten Missing Number Winter	.39	15%
Kindergarten Missing Number Spring	.03	0%
Kindergarten Missing Number Fall, Winter, Spring Average	.36	13%
Grade 1 Missing Number Fall	.43	19%
Grade 1 Missing Number Winter	.30	9%
Grade 1 Missing Number Spring	.32	11%
Grade 1 Missing Number Fall, Winter, Spring Average	.37	14%

Note. Four kindergarten students had missing fall scores.

Missing Number predictive validity correlations ranged from $\underline{r}$ = .03 in the spring kindergarten scores to .52 in the fall kindergarten scores. Therefore, beginning kindergarten scores were large predictors of Grade 3 MontCAS scores but decreased to medium in the winter and small in the spring. The mean of the three kindergarten scores indicated 13% explained variance on the grade 3 MontCAS scores. In grade 1 Missing Number scores were each medium predictors. The mean of the three first grade scores indicated 14% of the variance in the Missing Number variable explained its linear relationship with Grade 3 MontCAS scores. Considering these data, especially the mean of the 3 administrations would require one to reject the hypothesis (H_{04}) that there is no statistical relationship between MN and the grade 3 MontCAS.

Table 8

Correlations Between Average Fall, Winter, and Spring Scores for Each of the Measures

Variables	1	2	3	4	5	6	7	8	9	10	11
1 K OC	1.00	.60**	.71**	0.36	.77**	0.24	.48**	.47*	.68**	.53**	.44*
N	28	27	27	25	28	28	28	28	19	28	28
2 K NI	-	1.00	.67**	.48*	.40*	.70**	.55**	.59**	.81**	.62**	.62**
N		27	27	25	27	27	27	27	18	27	27
3 K QD	-	-	1.00	.48*	.38*	0.35	.52**	.49**	.58*	.55**	.59**
N			27	25	27	27	27	27	18	27	27
4 K MN	-	-	-	1.00	0.15	0.18	.42*	.43*	.70**	0.34	0.36
N				25	25	25	25	25	18	25	25
5 Gr1 OC	-	-	-	-	1.00	0.35	.52**	.42*	.60**	.41*	0.09
N					29	29	29	29	20	29	29
6 Gr1 NI	-	-	-	-	-	1.00	.64**	.52**	.56*	.58**	.47**
N						29	29	29	20	29	29
7 Gr1 QD	-	-	-	-	-	-	1.00	.79**	.72**	.74**	.39*
N							29	29	20	29	29
8 Gr1 MN	-	-	-	-	-	-	-	1.00	.69**	.70**	.370*
N								29	20	29	29
9 Gr1 MCBM	-	-	-	-	-	-	-	-	1.00	.86**	.57**
N									20	20	20
10 Gr2 MCBM	-	-	-	-	-	-	-	-	-	1.00	.49**
N										29	29
11 MONTCAS	-	-	-	-	-	-	-	-	-	-	1.00
N											29

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Note. N=29. K= kindergarten, G1 = grade 1, OC = Oral Counting, NI = Number Identification, QD = Quantity Discrimination, MN = Missing Number, GR1 MCBM = Grade 1 Mathematics Curriculum Based Measure, GR2 MCBM = Grade 2 Mathematics Curriculum Based Measures, and MontCAS = Grade 3 Montana Comprehensive Assessment System criterion test.

In general, the intercorrelations among the TEN measures were medium and large. The kindergarten TEN had large correlations with the same grade 1 TEN measure (i.e. OC and QD each had large correlations between same measures).

Each of the assessments showed significance to grade 3 MontCAS test scores except kindergarten Missing Number and grade 1 Oral Counting. Kindergarten NI, grade 1 NI, grade 1

QD and grade 1 MN showed significance with each of the other variables as well as the grade 3 MontCAS scores. Grades 1 and 2 M-CBM variables showed significance with all variables except kindergarten MN.

Question 5: Do Mathematics-Curriculum-Based Measurement in grades 1, and 2 and 3 predict student performance on third and fourth grade MontCAS?

The second cohort of 20 students did not have grade 3 M-CBM scores or grade 4 MontCAS scores and so only grades 1 and 2 M-CBM scores were correlated with grade 3 MontCAS scores from both cohorts for a total of 29 students.

Table 9

Correlation of Grade 1 and Grade 2 M-CBM with Grade 3 MontCAS Scores

Grade Level and Test Interval	Pearson r	Pearson r^2
Grade 1 M-CBM Fall	.25	7%
Grade 1 M-CBM Winter	.67	45%
Grade 1 M-CBM Spring	.47	22%
Grade 1 M-CBM Fall, Winter, Spring Average	.57	33%
Grade 2 M-CBM Fall	.41	17%
Grade 2 M-CBM Winter	.33	11%
Grade 2 M-CBM Spring	.59	34%
Grade 2 M-CBM Fall, Winter, Spring Average	.49	24%

Note. Nine grade 1 students had missing fall scores.

Mathematics Curriculum Based Measures' predictive validity correlations ranged from $r = .25$ in the fall of grade 1 to .67 in the winter of grade 1. Grade 1 fall scores indicated a small correlation then jumped to a large correlation in the winter and ended with a medium correlation for spring scores. The mean of the scores showed a 33% explained variance for Grade 3 MontCAS scores. The grade 2 scores showed medium correlations for fall and winter and a large correlation for grade 2 spring scores. Based on these findings, the mean of the scores provide 24% of the explained variance in the grade 1 M-CBM variable for Grade 3 MontCAS scores. Considering these data would require one to reject the hypothesis (H_{05}) which states there is no statistical relationship between M-CBM and the grade 3 MontCAS.

Question 6: Which of the Test of Early Numeracy (TEN) measures explains the most variance on the MontCAS and M-CBM assessments?

According to the previous findings for questions 1 through 4, Missing Number and Number Identification explained the most variance on the grade 3 MontCAS assessment. Overall, the kindergarten assessments explained more variance than the grade 1 assessments especially in Oral Counting where the grade 1 scores explained an average of 1% while the kindergarten scores were calculated to provide 20% explained variance (see Tables 4-7). Based on these findings, one would be required to reject the hypothesis (H_{06}) that each of the Tests of Early Numeracy measures would equally explain variance on grade 3 MontCAS.

To further investigate question 6, a sequential multiple regression procedure was conducted using SPSS software to determine which of the TEN measures along with the hierarchy of tests from kindergarten through grade 3 had the greatest explained variance for the grade 3 MontCAS scores and also to investigate the unique contribution of each variable with the overlapping effects of all other variables statistically removed (Pallant, 2007).

Four procedures were conducted; one procedure for each of the EM-CBM TEN measures beginning with Oral Counting and proceeding with Number Identification, Quantity

Discrimination, and Missing Number. The variables that were used for the calculation for the sequential regression were the averages of each student's test scores (fall, winter, and spring).

The variables were entered sequentially into hierarchical models in the order of the developmental progression of early mathematics assessments conducted in the district. Each procedure began with the kindergarten TEN averaged variable followed by the same grade 1 TEN averaged variable, then grade 1 M-CBM averaged variable, and finally the grade 2 M-CBM averaged variable. Each variable was entered into a new step of regression in the order that the students had been tested (kindergarten variables first and on through to grade 2). Each of the variables was assessed in terms of what it adds to the prediction of the MontCAS (dependent variable) after the previous variable or score.

Table 10 illustrates the results of the relative contribution of each of the kindergarten and grade 1 TEN measures along with the grade 1 and 2 M-CBM assessments in regard to the explained variance on the grade 3 MontCAS score.

Table 10

Variance in Grade 3 MontCAS scores Explained for the Mean of Each K-2 Formative Assessment

Formative Assessment Regression Equation	Sequential percent of variance in MontCAS explained for each formative assessment. K TEN Measure	 G1 TEN Measure given K TEN	 G1 CBM given K &1 TEN	 G2 CBM given K&1 TEN and G1 CBM
Oral Counting (OC)				
MontCAS=.44 (K OC) + 250	20%			
MontCAS = .93 (K OC) - .63 (G1 OC) + 310		36%		
MontCAS = .61 (K OC) - .75 (G1 OC) +.613 (G1 MCBM) + 309			55%	
MontCAS = .61 (K OC) -.78 (G1 OC) + .79 (G1 MCBM) - .18 (G2 MCBM) + 314				56%
Number Identification (NI)				
MontCAS=.62 (K NI) + 229	39%			
MontCAS = .57 (K NI) + .07 (G1 NI) + 224		39%		
MontCAS = .40 (K NI) + .07 (G1 NI) + .21 (G1 MCBM) + 223			40%	
MontCAS = .45 (K NI) + .04 (G1 NI) + .09 (G1 MCBM) + .11 (G2 MCBM) + 224				41%
Quantity Discrimination (QD)				
MontCAS=.59 (K QD) + 240	35%			
MontCAS = .49 (K QD) + .17 (G1 QD) + 236		37%		
MontCAS = .43 (K QD) - .19 (G1 QD) +.48 (G1 MCBM) + 309			44%	
MontCAS = .44 (K QD) -.19 (G1 QD) + .54 (G1 MCBM) - .08 (G2 MCBM) + 230				44%
Missing Number (MN)				
MontCAS=.36 (K MN) + 263	13%			
MontCAS = .25 (K MN) + .27 (G1 MN) + 247		19%		
MontCAS = -.10 (K MN) - .05 (G1 MN) +.68 (G1 MCBM) + 245			33%	
MontCAS = -.18 (K MN) - .03 (G1 MN) +.87(G1 MCBM) -.17(G2 MCBM) + 245				34%

Note. TEN = Tests of Early Numeracy, MontCAS = Montana State Criterion-Referenced Test, K = Kindergarten, G1 = Grade 1, MCBM = Mathematics-Curriculum Based Measures, G2 = Grade 2.

According to Table 10, each test from kindergarten to grade 1 increased the predictability of the grade 3 MontCAS scores. Grade 2 tests showed a plateau on the chart and a decrease indicated in the regression equation resulting in a 1% predictability increase from the grade 1 scores. The Missing Number test began at the lowest explained variance with 13% in kindergarten but continually rose through grade 1 and ended in Grade 2 with 34%, it also had the lowest predictability of the four outcomes. The Oral Counting test began with the next lowest variance in kindergarten but with the relative contributions of the tests through grade 2 it ended with the highest explained variance of 56%, which is almost three times the variance at the kindergarten level. The Number Identification and Quantity Discrimination had the most static results without much increase with Number Identification showing the least change indicating the kindergarten variance to be 39% and ending with the second grade CBM variance at 41%.

In summary, this chapter illustrated the results of the descriptive statistics, an overview of the correlations among the different data collected, and sequential regressions. The data used included kindergarten and first grade TEN measures, grades 1 and 2 M-CBM scores and lastly grade 3 MontCAS scores as the dependent variable.

Chapter 5

Discussion

Introduction

Global markets and technological advances have created an urgency for improving mathematics instruction and performance for United States students (Boaler, 2008; Jordan et al., 2010; Lago & DiPerna, 2010; NCTM, 2000). Based on national reports, barely one-third of U.S. fourth and eighth graders score at the proficient level on standardized tests (NAEP, 2009). These two facts raise concern for individual students in regard to the career choices available to them and for the overall society in maintaining a competitive place in the global market

Improving mathematics instruction is a fundamental approach to solving this problem and such instruction must begin with early intervention at the kindergarten and first grade level. Implementing formative assessment systems that track individual student progress and provide substance for the instructional decisions; is a crucial component of improving instruction (Clarke & Shinn, 2004). In addition to tracking student growth, such an approach allows the identification of students at the primary level who are at-risk of struggling in mathematics and can positively impact their chances of decreasing the performance gap between them and their higher achieving peers (Gersten, Jordan, & Flogo, 2005; Kashi, 2008).

During the past 50 years, primary reading approaches aimed at helping struggling readers has been the main focus of intervention research for early education. In doing so, critical basic skills of phonological awareness have been identified in this area (Thurber, Shinn, & Smokowski, 2002). Recently, the same movement in research is identifying the critical mathematics basic skills that provide a solid foundation for mathematical understanding in kindergarten and first grade students. Recent research has explored the area of number sense which begins development in infancy and progresses throughout one's life; but, especially critical at the primary level in building a strong foundation of flexibility and fluency with numbers (Jordan et al., 2009; VanDerHeyden, 2010).

This chapter brings forth the knowledge gained from this correlational research, subsequent recommendations, and implications for further research. Correlation coefficients were calculated between formative assessments (TEN and M-CBM) and a criterion assessment (MontCAS) in order to determine the existence and/or strength of relationships. The determination was made by statistical calculations, which investigated whether TEN and M-CBM scores within a certain range were related to a certain range of the grade 3 MontCAS scores.

Research Questions' Interpretations of Findings

Do Tests of Early Numeracy (TEN) in kindergarten and grade 1 correlate with mathematics performance in grade 3? TEN assessments included Oral Counting (OC), Number Identification (NI), Quantity Discrimination (QD), and Missing Number (MN). M-CBM assessed computational fluency. The tests were conducted three times throughout each school year approximately 13 weeks apart in the fall, winter, and spring of kindergarten and first grade.

One of the main purposes of the current study was to explore the predictive validity of kindergarten and grade 1 TEN scores relative to grade 3 MontCAS scores. The outcome of this component was a two-part examination of the TEN by the highest predictability broken down by each TEN skill (OC, NI, QD, and MN) variable and secondly by grade level.

The first analysis findings indicated that virtually all four of the TEN variables had some degree of predictability for educational achievement as defined by the grade 3 MontCAS.

Kindergarten Number Identification (NI) and Quantity Discrimination (QD) had the most explained variance followed by kindergarten Oral Counting (OC).

Based on these findings, these three variables could be used to identify primary grade students who are not-at-risk of struggling in mathematics and/or may perform at Proficient or Advanced levels on the grade 3 MontCAS.

The examination of variables by grade level indicated that the kindergarten scores had the highest predictability especially in NI, QD, and OC. The first grade levels showed half as much of the kindergarten explained variance with grade 3 MontCAS. Kindergarten OC indicated explained variance but grade 1 did not.

The second analysis looked for the difference in predictability between kindergarten and grade 1 scores on the TEN skills. Based on these findings, the kindergarten TEN measures showed stronger predictability for identifying students who may score at proficient or above levels on the grade 3 MontCAS as compared to the first grade measures.

Table 11

Comparison of Predictive Validity Results of TEN Longitudinal Studies

Study	**n**	**Length of Study**	**Grade Level**	**Independent Variables**	**Dependent Variables**	**Results Strongest to Least Predictive Validity**
Baglici (2008)	61	2 year	K-1	OC, NI, QD, MN, VQD	Gr1 M-CBM, ODR, ACES	MN, NI, OC, QD
Burland (2011)	29	4 year	K-3	K & Gr 1: OC, NI, QD, MN; Gr.1 and 2: M-CBM	Gr3 MontCAS	NI, QD, MN, OC K scores were stronger than G1 &2
Chard et al., (2005)	168 (K) 207 (G1)	1 year	K-1	Counting, NW, NI, QD, MN	NKT	MN, QD, NI, NW, Counting
Clarke & Shinn (2004)	52	1 year	Gr 1	OC, NI, QD, MN	M-CBM, WJ-AP, NKT	QD, MN, NI, OC
Lembke et al., (2008)	77 (K) 30 (G1)	1 year	K-1	NI, QD, MN	TR, SESAT	QD, NI, MN
Lembke & Foegen (2009)	384 (K&G1)	1 year	K-1	QD, QA, MN, NI	TR, SESAT	MN, NI, QD, QA

Note. OC = Oral Counting, NI = Number Identification, QD = Quantity Discrimination, MN = Missing Number, VQD = Visual Quantity Discrimination, M-CBM=Mathematics Curriculum Based Measurement, ODR = Office Discipline Referrals, ACES = Total Mathematics Score on the Academic Competency Evaluation Scales, WJ-AP = Woodcock Johnson-Applied Problems, NKT = Number Knowledge Test, Counting = Assessments that include counting: to 20, from 6 and on, from 3 and on, by 10s, by 5s, by 2s; NW = Number Writing, TR = Teacher Ratings, SESAT = Stanford Early School Achievement Test, QA = Quantity Array.

Clarke and Shinn (2004) studied TEN measures as independent variables with performance on Walcott Johnson - Applied Problems (WJ-AP) as the dependent variable and found QD to have the highest median correlation followed by MN then NI and finally OC. Each demonstrated strong relationships; therefore, showing predictability (Clarke & Shinn, 2004).

Charde et al. (2005) replicated the Clarke and Shinn (2004) predictive study using TEN measures as the independent variables and the Number Knowledge Test (NKT) as the dependent variable. The predictability measures repeated the Clarke and Shinn (2004) results. Lemke and Foegen (2009) also studied TEN measures and found parallel results for predictive validity for NI, QD, and MN tasks (they did not include OC in their study).

Did Mathematics-Curriculum Based Measurement in grades 1, 2, and 3 correlate with performance on third and fourth grade MontCAS? Grades 1 and 2 Mathematics Curriculum Based Measure (M-CBM) variables were examined for explained variance of educational achievement as defined by grade 3 MontCAS scores. This investigation of predictability was broken into two different analyses. The first analysis was done by looking for the highest explained variance among the fall, winter, and spring scores. The second analysis looked at the difference in the grade level scores to see which grade level indicated higher explained variance.

In general, all three of the administrations of the M-CBM measures (fall, winter, and spring) in both grades 1 and 2 showed predictability of grade 3 MontCAS scores. In grade 1 the winter score and in grade 2 the spring scores demonstrated the largest correlations. Based solely on the grade 1 scores, the winter score demonstrated the highest predictability with a large correlation of .67 as compared to the fall score, which demonstrated the least predictability with a small correlation of .25. The spring second grade scores showed a large correlation with .59 while the other two scores were in the medium range; winter scores showed the lowest with .33 followed by .41 in the fall.

When comparing the first grade to the second grade correlations, first grade scores showed higher predictability with a large correlation of .57. The second grade demonstrated a medium correlation of .49. Based on these findings, grade 1 and 2 M-CBM scores could also be used as an indicator to identify primary grade students who are achieving in mathematics and/or who may score at the proficient level or above on the grade 3 MontCAS.

Which Tests of Early Numeracy (TEN) measures explained the most variance on the MontCAS and M-CBM assessments?

The last investigation of the current study looked at the possibility that using multiple predictors would improve predictability and/or explained variance. A sequential regression procedure was used to calculate the explained variance of the student assessments taken from kindergarten through grade 2 as defined by the grade 3 MontCAS scores. Four procedures were entered each starting with an individual TEN measure and proceeding with the next grade level test. The test variables used were the average of the fall, winter, and spring scores. The analysis was broken down into two analyses first by each TEN variable and a second analysis by grade level.

The first analysis found that Oral Counting (OC) demonstrated the most linear growth in the sequential regression investigation over the three-year period of testing. The other three TEN variables produced an even distribution. The second analysis, which compared explained variance by grade levels indicated that the predictability from first to second grade M-CBM scores nearly stayed the same. Based on these findings, using the TEN and grade 1 M-CBM scores as multiple predictors showed nearly 50% predictability or explained variance of the performance on grade 3 MontCAS. In short, the second grade M-CBM scores did not add to the first and second grade sequential regression predictability of performance on the grade 3 MontCAS; therefore did not contribute to the predictability of the criterion variable in this study. Still neither this research nor other similar research designs would be logically capable of concluding that the curriculum components that generated those scores should be discontinued based on these results. This was a correlational study not a cause and effect study. Just as a strong correlation does not constitute proof of causality, neither does the lack of a correlation mean there is no causality; to do so would commit the logical error of *accepting* the null hypothesis (Gay, Mills, & Airasian, 2009).

Similar research by Lembke et al. (2008) was conducted using a multiple regression analysis. This study applied a two-level hierarchical linear growth analysis to measure student progress over time for MN, NI, and QD. Tests were given monthly (instead of 3 times each year) and the results indicated that both kindergarten and first grade students showed a significant linear growth in NI and curvilinear in QD and MN. Lemke et al. (2008) suggested that the QD and MN measures either indicate student learning happens in bursts or QD and MN are not good indicators.

Other findings. The longitudinal outcome delineated by descriptive statistics for each assessment indicated plateaus or drastic decreases in the mean of the scores between spring and fall of adjacent years. For example, the Oral Counting (OC) and Quantity Discrimination (QD) measures between kindergarten and first grade stayed the same and then continued to rise throughout the end of the first grade TEN assessments. However, Number Identification (NI), Missing Number (MN), and Math Curriculum Based Measures (M-CBM) fall scores showed decreases of as much as fifty percent from previous spring scores. Although NI and M-CBM scores had surpassed the fall entry score by the winter administering of the test, MN had merely broken even which indicates that it took nearly one-third of the school year to bring those skills back to the previous year's spring level.

This is an important practical observation that research has cited as contributing to the patterns of educational stratification especially between children from different socioeconomic backgrounds (Alexander, Entwisle, & Olson, 2007). Decreased scores from spring to fall indicate another reason to utilize formative testing as an RTI approach especially to students at-risk of struggling in mathematics. Students who enter kindergarten and first grade with a less-developed sense of number and number relations would have a better chance of being identified early and given individualized assistance based on the demonstrated gaps in mathematical understanding. The gap in summer learning between students from lower and higher family socioeconomic levels consistently widens between primary and high school and creates lifelong effects (Alexander, Entwisle, & Olson, 2007).

Implications for Practice

The results of this study can demonstrate ways to identify primary grade students who may be at-risk of struggling as they progress through grade levels. Principals, teachers, and other instructional supervisors can use this information to justify the use of Response to Intervention (RTI) approaches that provide early intervention and prevent severe mathematics difficulties from developing. Using scientifically-based researched formative assessment systems as part of an RTI approach can identify gaps in mathematical understanding through the longitudinal data produced. This data can guide instructional decisions before students fall farther and farther behind (Bryant & Bryant, 2008; Stecker, Fuchs, & Fuchs, 2008).

When summative standardized tests such as MontCAS alone are used to assess student growth it is not possible to use the results to inform instruction. The test results only tell how students perform at the end of the year and the results become available only after students have departed for the summer break. With RTI approaches the results of the longitudinal ratio level data collected throughout the year allows teachers to make data-driven instructional decisions. One of the most effective components of the approach is the ability for teachers to track student growth and individualize instruction especially for primary students identified as mathematically at-risk as they progress (Bryant et al., 2008; Gersten et al., 2008).

In a practical sense, the amount of time spent administering the RTI formative assessments in this study amounted to about two days (one day per semester/half-a-day per quarter) for grades kindergarten and first grade based on a classroom of 30 students. The TEN assessments are administered individually to students so the time taken is greater than the M-CBM assessments. Still, this may or may not be time taken away from instruction depending upon a teacher's organization and management of classroom activities. Such activities may include using academic stations for independent learning while a teacher administers to individual students. The first and second grade M-CBM assessments take 30 minutes from the school year to administer. Each test takes a maximum of 10 minutes to give to the whole class at one time. Administering the M-CBM during the three 13-week periods in the fall, winter, and spring uses a maximum of 30 minutes throughout the school year. Considering the small amount of time taken to administer an RTI system the study indicates that the information it provides for designing instruction is time efficient and worthwhile.

Analyzing the difference in predictability between grades with multiple variables in the sequential regression analysis, grades kindergarten and first grade scores showed higher predictability than the second grade scores. In fact, when the second grade scores were included in the sequence, they did not add to the predictability of performance on the grade 3 MontCAS. This is not to say that grade 2 formative assessments are not valuable, they just indicated a lower correlation for grade 3 performance on the MontCAS. This does not mean that the data generated from use would not inform instruction and assist in tracking student growth.

These findings echo research on struggling students by suggesting that there is a correlation between kindergarten and first grade children's critical number sense skills and future mathematical achievement. Even though this study analyzed students who performed at Proficient and Advanced levels on a criterion referenced test in grade 3, moderate correlation between the early mathematics test scores and their 3rd grade performance showed the same strength in correlations. With this in mind, this study concurs with other research that RTI services allow gaps in student learning to be identified and intervened early so that at-risk students may be prevented of failing to keep up with higher achieving peers (Bryant et al., 2008, Gersten et al., 2008).

Weaknesses of the Study/Missing Data Pieces

The limited generalizability of the study is a weakness. This weakness is two-fold including the longitudinal focus and the size of the district studied. In regard to the longitudinal focus, as with any study spanning time, family dynamics change in many ways from the make-up of the family to the career and job changes that require students to leave the participating school district (Baglici, 2008). In regard to size of the district, the participating school was small and rural so two cohorts were combined for statistical purposes. Furthermore, with any span of time in a district striving to improve, changes are made in the curriculum continually. For this reason, the participating school changed from M-CBM assessments in the third grade year of cohort 2's third grade year (year 4 of the study). This decreased the data available and also made it impossible to investigate concurrent validity of the M-CBM in relation to the grade 3 MontCAS.

Another weakness was that only student scores were included in the available data; there was no demographic information about the population. A larger population and a database that included gender and socioeconomic levels may have produced more comprehensive results. Alexander, Entiwisle, and Olson (2007) found that socio-economic levels impact student mathematical achievement. Attendance and behavioral referral data may have also been useful.

Solely using MontCAS scaled scores may have posed a limitation or weakness because they are made up of unequal intervals and have no absolute zero. These were the only scores available from the district and may or may not have been less accurate when analyzing correlations with TEN and M-CBM.

Implications for further research

To increase the scope of this research, the following implications are offered. Early mathematics research should continue to strive to learn more about indicators that predict later mathematics performance to improve instructional practices so that students do not fall behind peers and benchmark performance levels. Such studies should include a more diverse and larger population to include the full spectrum of MontCAS scores. For example, none of the students in this study performed at the lowest level of the MontCAS but correlations that were found in other studies including at-risk students showed the same strength of predictive validity. Studying populations with a more diverse population that has students at every level would allow for another component of research based on interventions taking place and the impact on student learning from those interventions.

It would be beneficial for future correlational studies specific to Montana to use raw data from the MontCAS rather than only scaled scores. In addition, it would be of interest to compare the results of both (raw and scaled scores) to determine which scores were more accurate for predictability measures. Scaled scores statistically responded to only the middle scores on the MontCAS because of the uneven interval distribution with no absolute zero and a capped high score used for a large proportion of the scaled scores (see Table 2).

The ideal longitudinal correlational study would be one that would take population data from the whole state collected from all schools using RTI approaches and correlating them with the MontCAS at all grade levels. The results could be used to provide more evidence for the use of TEN measures or other EM-CBM assessments as a tool for identifying students in need of intervention and whether the intervention is needed because of learning disabilities or inadequate instruction (Baglici, 2008).

Finally, studies on retention approaches for avoiding students' drop in achievement between spring and fall of adjacent school years would be beneficial to districts that struggle to make AYP due to low student achievement on high stakes tests. Future research might investigate the impact of RTI implementation on summer programs and whether this decreases the gap between low level and high level achievement of students.

Conclusion

In conclusion, the results of this study paralleled previous studies that found that TEN scores demonstrate implications of use as early indicators of mathematics skills. These studies also included test-retest, reliability, concurrent validity, content validity, and predictive validity over one and two years. The current study looked mainly at predictive validity by extending the prior research longitudinally to a four-year span. The results can be used to provide insight to teachers, principals, and other instructional supervisors on improving instruction through the use of RTI and implementing early intervention to primary grade students who are at-risk to struggling in mathematics.

As in other studies conducted, varying degrees of predictability was found for the Montana state level achievement test. Some levels of predictability were strong enough to suggest additional research is warranted. Should a much larger scale of research be conducted with acceptable levels of predictability found, then Montana educators would have a very useful and simple tool to predict students' performance on standardized assessments *a priori* and provide a potentially effective opportunity to individually investigate factors that may be hindering achievement.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167-180.

Baglici, S. P. (2008). The long-term predictive validity of early mathematics curriculum-based measurement (Doctoral dissertation). Available from ProQuest database (UMI No. 3325449).

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *Journal of Special Education, 38*(2), 66-79.

Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*, 69-79.

Bass, H. (2003). Computational fluency, algorithms, and mathematics proficiency: One mathematician's perspective. *Teaching Children Mathematics, 9*(6), 322-327.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*(4), 333-339.

Boaler, J. (2008). *What's math got to do with it? How parents and teachers can help children learn to love their least favorite subject.* New York, New York: Penguin Books LTD.

Boerst, T. A., & Schielack, J. F. (2003). Toward understanding of "computational fluency". *Teaching Children Mathematics*, *9*(6), 292-293.

Bryant, B. R., & Bryant, D. P. (2008). Introduction to the special series: Mathematics and learning disabilities. *Learning Disability Quarterly*, 3-8.

Bryant, D. P., Bryant, B. R., Gersten, R. M., Scammacca, N. N., Funk, C., Winter, A., et al. (2008). The effects of tier 2 intervention on the mathematics performance of first grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly, 31*, 47-63.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention*, *30*(2), 3-14.

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

Christ, T. J., & Vining, O. (2006). Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. *School Psychology Review, 35*(3), 387-400.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Crawford, L., & Ketterlin-Geller, L. R. (2008). Improving math programming for students at risk. *Remedial and Special Education, 29*(1), 5-7.

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

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493-513.

Gay, L., Mills, G. E., & Airasian, P. (2009). *Educational Research Competinecies for Analysis and Applications* (9th Edition ed.). Upper Saddle River, New Jersey: Pearson.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*(1), 18-28.

Gersten, R., & Jordan, N. C. (2005). Early screening and intervention in mathematics difficulties: The need for action. *Journal of Learning Disabilities, 38*(4), 291-292.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*(4), 293-304.

Gilbertson, D., Witt, J., Duhon, G., & Dufrene, B. (2008). Using brief assessments to select math fluency and on-task behavior intervention of treatment utility. *Education and Treatment of Children, 31*(2), 167-181.

Griffin, S. (2003). Laying the foundation for computational fluency in early childhood. *Teaching Children Mathematics, 9*(6), 306-309.

Gurganus, S. (2004). 20 ways to promote number sense. *Intervention in School and Clinic, 40*(1), 55-58.

Hoover, J. J., & Love, E. (2010). Supporting school-based response to intervention: A practitioner's guide. *Teaching Exceptional Children, 43*(3), 40-48.

Jiban, C. L., & Deno, S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance: Are simple one-minute measures technically adequate? *Assessment for Effective Intervention, 32*(2), 78-89.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Reveiw, 39*(2), 181-195.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850-867.

Kamii, C. (2000). *Young children reinvent arithmetic: Implications of Piaget's theory.* New York, New York: Teacher's College.

Kashi, T. L. (2008). Response to intervention as a suggested generalized approach to improving minority AYP scores. *Rural Special Education Quarterly*, 37-44.

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

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher, 34*(8), 3-13.

Lago, R. M., & DiPerna, J. C. (2010). Number sense in kindergarten: A factor-analytic study of the construct. *School Psychology Review, 39*(2), 164-180.

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research and Practice, 24*(1), 12-20.

Lembke, E., Foegen, A., Whittaker, T. A., & Hampton, D. (2008). Establishing technically adequate measures of progress in early numeracy. *Assessment for Effective Intervention, 33*(4), 206-214.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, 451-459.

Montana Office of Public Instruction. (2010, March 1). *CRT and CRT Alt Test Administration.* Retrieved July 19, 2010, from Montana Office of Public Instruction: http://opi.mt.gov/curriculum/MontCAS/#gpm1_7

National Assessment of Educational Progress. (2009). *The Nation's Report Card Mathematics 2009.* U.S. Department of Education. Washington. DC: Institute of Education Sciences.

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

Pallant, J. (2007). *SPSS Survival Manual* (3rd Edition ed.). New York: McGraw Hill Open University Press.

Pearson Education, Inc. (2010, July 1). *AIMSweb.com Benchmark and progress monitoring for grades K-8.* Retrieved September 2, 2010, from AIMSweb Assessment and Data Management for RTI: http://www.aimsweb.com/news/39/42/Learning-Disabilities-and-Early-Intervention-Strategies/d,Articles/

Reys, B., & & Lappan, G. (2007). Consensus or confusion? The intended math curriculum in state-level standards. *Phi Delta Kappan, 88*(9), 676-680.

Reys, R. E. (1998). Computation versus number sense. *Mathematics Teaching in the Middle School, 4*(2), 110-12.

Russell, S. J. (2000). Developing computational fluency with whole numbers. *Teaching Children Mathematics, 7*(3), 154-158.

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*(1), 37-59.

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

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. Shinn, *Advanced Applications of Curriculum-Based Measurement* (pp. 1-31). New York: Guilford Press.

Stecker, P. M., Fuchs, D., & Fuchs, L. S. (2008). Progress monitoring as essential practice within response to intervention. *Rural Special Education Quarterly, 27*(4), 10-17.

Thurber, R. S., Shinn, M. R., & Smokowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review*, 498-512.

U.S. Department of Education. (2002, January 8). *Ed.gov U.S. Department of Education: Elementary and Secondary Education Act.* Retrieved October 11, 2009, from http://www.ed.gov/policy/elsec/leg/esea02/index.html

United States Department of Education. (2004, September 15). *Elementary and Secondary Education Act Title IX General Provisions*. Retrieved July 31, 2010, from Ed.Gov U.S. Department of Education: http://www2.ed.gov/policy/elsec/leg/esea02/pg107.html#sec9101

Van De Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics teaching developmentally* (7th Edition ed.). Boston, MA: Pearson Education, Inc.

VanDerHeyden, A. M. (2010). Determining early mathematical risk: Ideas for extending the research. *School Psychology Review*, *39*(2), 196-202.

Appendix A

List of Acronyms

AYP – Adequate Yearly Progress

DSTP – Delaware Student Testing Program

EM – Early Mathematics

EM-CBM – Early Mathematics Curriculum Based Measure

ESEA – Elementary and Secondary Educational Act

ESL – English as a Second Language

IDEA – Individuals with Disabilities Act

KNSB – Kindergarten Number Sense Battery

M-CBM – Mathematics –Curriculum Based Measure

MD – Mathematics Difficulties

MN – Missing Number

MontCAS – Montana Comprehensive Assessment System

NAEP – National Assessment of Educational Progress

NCLB – No Child Left Behind

NCTM – National Council of Teachers of Mathematics

NI – Number Identification

NKT – Number Knowledge Test

NSB – Number Sense Brief

NST – Number Sense Test

OC – Oral Counting

PM-CBM – Preschool Mathematics Curriculum-Based Measurement

PSI – Personalized System of Instruction

QD – Quantity Discrimination

RTI – Response to Intervention

SESAT – Stanford Early School Achievement Test

SRB – Scientifically Research Based

TEN – Test of Early Numeracy

TEMA – Test of Early Mathematics Ability

WJ-AP – Woodcock Johnson-Applied Problems

Appendix B

TEN and MontCAS Scatterplots

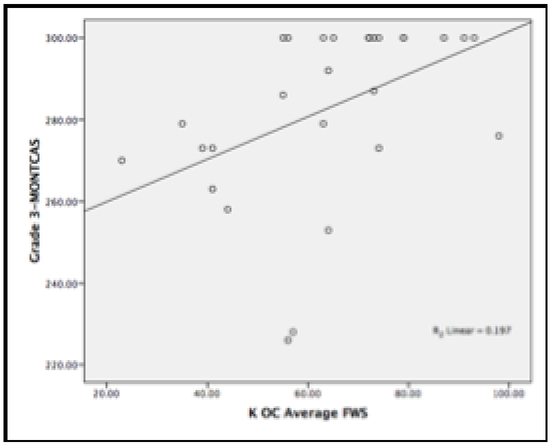


Figure 9. Kindergarten Oral Counting Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The kindergarten Oral Counting association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Oral Counting scores is .44 (see Table 4) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

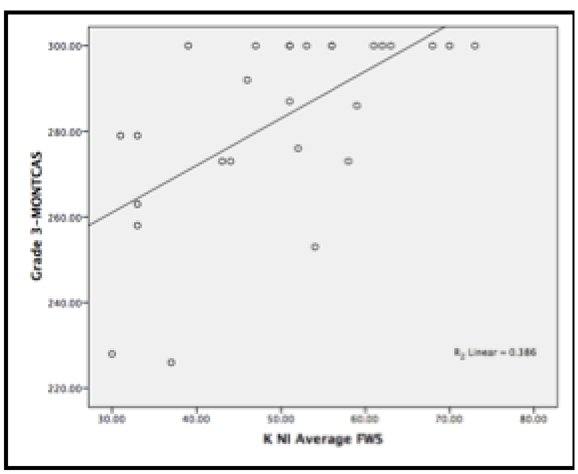


Figure 10. Kindergarten Number Identification Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The kindergarten Number Identification association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Oral Counting scores is .62 (see Table 5) which is a strong correlation. The outliers are in the y-direction or vertical direction.

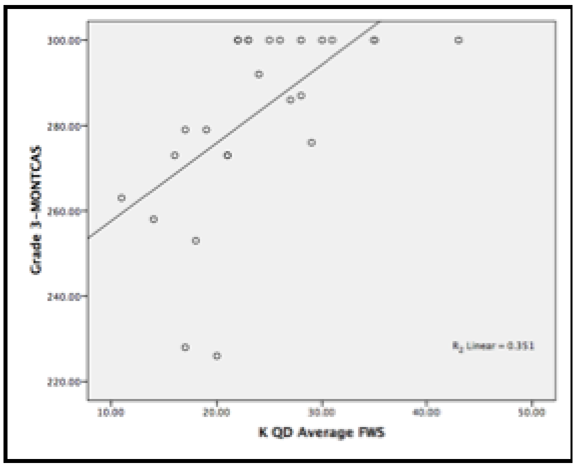


Figure 11. Kindergarten Quantity Discrimination Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The kindergarten Quantity Discrimination association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Oral Counting scores is .59 (see Table 6) which is a strong correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

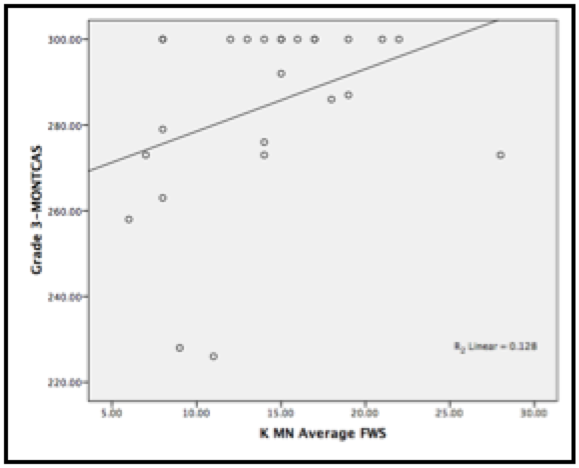


Figure 12. Kindergarten Missing Number Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS.

The kindergarten Missing Number association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Oral Counting scores is .36 (see Table 6) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

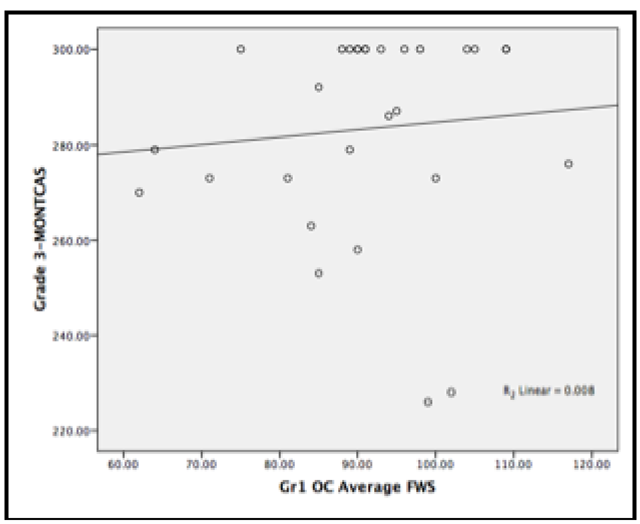


Figure 13. Grade 1 Oral Counting Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS.

The grade 1 Oral Counting association with the grade 3 MontCAS score does not show an association in the linear form. The r-value for the mean of the kindergarten Oral Counting scores is .09 (see Table 4) which corresponds to the linear regression line which is shows neither a positive nor negative correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

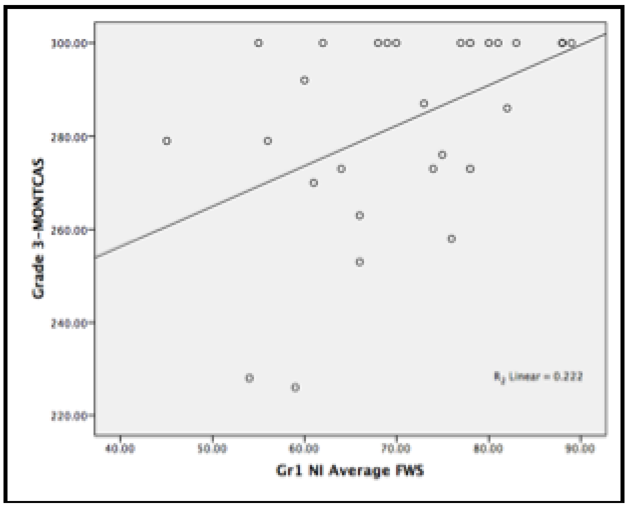


Figure 14. Grade 1 Number Identification Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The Grade 1 Number Identification association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Number Identification scores is .47 (see Table 5) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

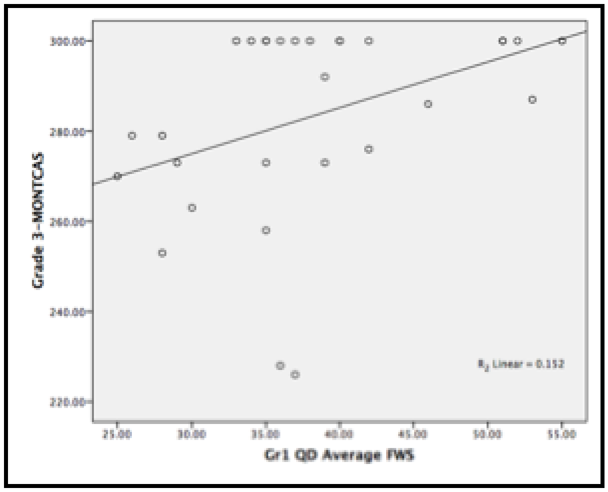


Figure 15. Grade 1 Quantity Discrimination Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The Grade 1 Quantity Discrimination association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Number Identification scores is .39 (see Table 6) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

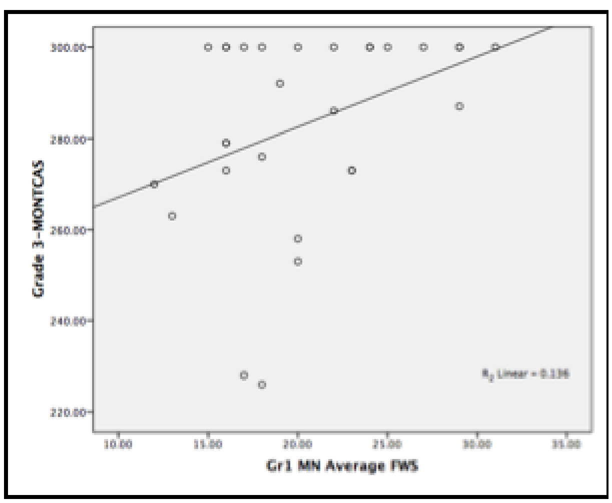


Figure 16. Grade 1 Missing Number Mean from fall, winter, and spring scores correlated with Grade 3 MontCAS

The Grade 1 Missing Number association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Number Identification scores is .37 (see Table 7) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

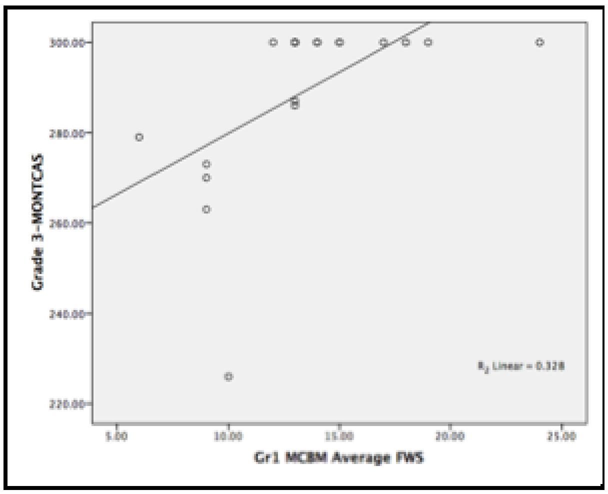


Figure 17. Grade 1 M-CBM Mean from fall, winter, and spring scores correlated with grade 3 MontCAS.

The Grade 1 M-CBM association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Number Identification scores is .57 (see Table 9) which is a strong correlation. The outlier is in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).

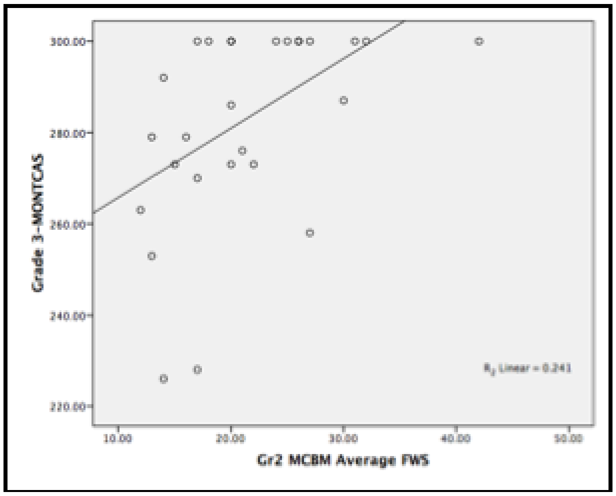


Figure 18. Grade 2 M-CBM Mean from fall, winter, and spring scores correlated with grade 3 MontCAS.

The Grade 2 M-CBM association with the grade 3 MontCAS score has a positive association in the linear form. The r-value for the mean of the kindergarten Number Identification scores is .49 (see Table 9) which is a moderate correlation. The outliers are in the y-direction or vertical direction showing the two students who performed in the Nearing Proficient Level of the MontCAS (see Figure 4).