TWO FORMS OF MATH CURRICULUM-BASED MEASUREMENT: AN EXAMINATION OF PREDICTIVE VALIDITY AND TEACHER ACCEPTABILITY

A Dissertation
Submitted to
The Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Bridget S. Sweeney
May, 2009

©

Copyright

2009

by

Bridget S. Sweeney

All Rights Reserved

ABSTRACT

Two forms of math curriculum-based measurement: An examination of predictive validity and teacher acceptability

Bridget S. Sweeney
Doctor of Philosophy
Temple University, 2009
Doctoral Advisory Committee Chair: James E. Connell

Curriculum-based measurement (CBM) is a fast, reliable, and valid set of procedures for measuring student progress in basic skills. However, the predictive validity of math computations and applications probes administered three times across the year (fall, winter, and spring) has not been compared to the TerraNova standardized assessment, nor has CBM's acceptability from corrective education teachers serving students out of the classroom been explored. In this study, nine corrective education teachers working in a large mid-Atlantic urban school district and providing corrective education services to private and parochial schools participated by administering math computations and concepts and applications CBM probes to second and/or third grade students three times. There were 453 second grade students in 19 nonpublic schools and 371 third grade students in 16 nonpublic schools who completed three computations and applications CBM probes at three different points in the year. Of these students, 133 second grade students in 13 parochial schools as well as 108 third grade students in 12 parochial schools completed the TerraNova math assessment. Through the use of multiple regression analyses, it was determined that the concepts and applications probes had significant levels of predictive validity while the computations probes were not found to have any predictive validity when compared to the Normal Curve Equivalent of the Total Math subtests. The nine

corrective education teachers also completed three versions of the Assessment Rating Profile-15 (ARP-15), one prior to the use of CBM and two following its use. No significant difference was identified when comparing the results of the first rating scale with either of the second or when comparing the two different rating scales administered at the end of the year. Overall, the concepts and applications probes were found to be significantly predictive of performance on the TerraNova, Second Edition, Total Math subtests, despite the fact that the teachers did not indicate a strong like or dislike towards the use of CBM tools.

ACKNOWLEDGMENTS

I am forever grateful to James E. Connell who was a blessing to me on this journey. His support, encouragement, and positive spirit made me realize that graduating would become a reality. His confidence in me pushed my limits and gave me confidence in myself and my abilities. I am a better school psychologist after having had his guidance and expertise along this journey.

I appreciate the assistance and guidance of Catherine Fiorello and Joseph Ducette, members of the committee who made this dissertation what it is today.

I am certain that without the guidance of Jean Boyer, this dissertation would have never been written. Jean's guidance has led me down the path that led to the accomplishment of this goal. I appreciate the consultation she provided to me about my life, and her efforts to intervene in order to create the best possible educational experience for me.

I am thankful to Erin and Joe who were my first learning partners. They know me better than anyone, and they are not afraid to let me know that. I know that they will always be there for me, as they were through all of my years of education, and I hope they know that I will forever be there for them.

I must acknowledge all of the friends I made while at Temple University, especially Kristy, Tiffany, Beth, Denise, and Jen who gave me the daily support and encouragement to make this possible. Honestly, I would not have survived this process without their influences, and most of all without their notes and homework.

I would like to acknowledge Mike, Suzanne, Ellen, Deborah, Dale, and everyone from Elwyn who allowed me the opportunity to grow as a school psychologist, as well as give me the gift of completing my dissertation. My time spent at Elwyn has been important to my development in the field and as a person. In addition, their assistance with data collection literally has made this degree a possibility.

I am grateful for all of the encouragement and support provided to me by the people at Father Judge High School. Without the understanding provided to me by the people here, I would have not been able to complete my goal.

I am thankful to everyone from St. Basil Academy and DeSales University who provided me with the education and work ethic foundations to allow me to continue my education until today and beyond. I am grateful to my many friends from these institutions who made my time at these places so much more enjoyable.

To all of the Slavin and Sweeney families, I would like to extend special thanks. As I have experienced throughout my life, no matter the type of support I needed these past years, they were there to provide it. They each deserve a piece of this honor because they all played a role in my development that helped me achieve this goal. I am proud to be a member of these families because they represent the ideals that I live every day: hard work, commitment, faith, compassion, dedication, and loyalty.

I am very grateful to my best friend and future life partner, Michael. He not only provided me with the desk where I spent many a night writing this tome, he gave me all of his

emotional support and encouragement that afforded me the strength to carry on every day. He walked through this journey with me, and his sacrifices have allowed me to pursue my dream. I am looking forward to creating and fulfilling new dreams as we continue to walk through our journey together for the rest of our lives.

DEDICATION

The people in this world that I most admire the most are the people to whom I would like to dedicate my doctorate. I am dedicating my dissertation and my degree to my first teachers, Joseph and Sarita Sweeney: this is not only *for* them, it is *because* of them. From my first day of pre-school until this current graduation day, I have constantly felt their undying support and encouragement towards my education and all of my other life endeavors. They have stood by me through the challenging times and celebrated with me during the momentous times. I know that they personally experienced it all with me, and they deserve to share this honor with me. However, without their strength, support, and encouragement, I would not have been able to accomplish my ultimate goal. It has been said to me that "the apple does not fall far from the tree." I am honored and blessed to be a part of their tree. I am proud to be their daughter, and I am proud to have been the one to bestow upon them their letters of honor, "MOM" and "DAD." A bhuí le máthair agus athair.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

Throughout history, tests have been used to make decisions about the aptitude and functioning level of individuals (Kaufman, 2000). More recently, student performance on tests of academic achievement has been used to rate the success of a school district and ultimately determine a school district's accountability towards meeting the educational needs of students (Braden, 2007; Hintze & Silberglitt, 2005; Jitendra, Sczesniak, & Deatline-Buchman, 2005; Katsiyannis, Zhang, Ryan, & Jones, 2007; McGlinchey & Hixson, 2004; No Child Left Behind Act of 2001 [NCLB]; Silberglitt, Burns, Madyun, & Lail, 2006). It is because of their growing importance that preparation for these assessments has become paramount to the academic success of students.

Within the public education system, tests have been used for many purposes. For example, Intelligence Quotient (IQ) batteries and achievement tests have been used to determine whether a discrepancy exists between a student's ability (i.e., performance on intelligence test) and achievement (i.e., performance on tests of academic achievement) in order to determine a student's Special Education eligibility (Francis et al., 2005; Fuchs & Fuchs, 2007; Fuchs, Fuchs, & Zumeta, 2008; Hilton, 2007; McCardle, Keller-Allen, & Shuy, 2008; Peterson & Shinn, 2002). This approach to eligibility determination, called the IQ/Achievement Discrepancy Model has received increasing criticism in the past 20 years (Francis et al., 2005; Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klingner, 2007; McCardle et al., 2008; Willis & Dumont, 2006). Critics have questioned the use of the Discrepancy Model for reasons including the validity of the scores obtained from intelligence tests as an index of overall ability (Francis et al., 2005; McCardle et al., 2008), the diagnostic criteria used (Francis et al., 2005; NASDE, 2005) and the limited population samples used to establish the normed data (Harry & Klingner, 2007). Claims of racial and ethnic bias of the tests have been made as well as concerns for the psychometric properties of these tests, including reliability and validity (Kaufman, 2000; Harry & Klinger, 2007; Marston, 1989; McCardle et al., 2008; Sattler, 2001). In addition, many have doubted the reliability and validity of the discrepancy component noting that the formula used to determine the discrepancy is not consistent among school districts, with some places using a regression-based approach, others specifying defining a discrepancy as deviation from the mean, and still others using simple subtraction to determine the difference between the two test scores (Francis et al., 2005; Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klinger, 2007; McCardle et al., 2008; Willis & Dumont, 2006).

The discrepancy model has been termed the “wait to fail” model by critics because students must fall behind in curriculum before being identified. Thus, early intervention and preventative measures are often too late (Fuchs & Fuchs, 2007; Fuchs et al., 2008; McCardle et al., 2008). However, others found merit within the framework of the discrepancy model for identification and eligibility determination of students with learning disabilities.

There has been evidence that the information obtained through these assessments can be useful in identifying students with special learning needs; however, these assessments are best suited when used with other measures and procedures in order to conduct a comprehensive evaluation (Fiorello, Hale, & Snyder, 2006; Hale, Kaufman, Naglieri, & Kavale, 2006; Naglieri & Kaufman, 2008; Willis & Dumont, 2006). Recent additions to The Individuals with Disabilities Education Improvement Act of 2004 (IDEIA) indicated an alternative procedure that can be used in identifying students eligible for special education. Specifically, IDEIA states that the discrepancy model does not need to be the only means to decide whether a student is eligible for special education services as a student with a learning disability (Fuchs & Fuchs, 2007; IDEIA, 2004; Zirkel, 2007). As an alternative, IDEIA states that the Response-to-Intervention (RtI), process can be used to determine eligibility for special education services, as decided by the school district (Ardoin, 2006; Barnett, VanDerHeyden, & Witt, 2007; Daly, Martens, Barnett, Witt, & Olson, 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Hilton, 2007; Shinn, 2007; Willis & Dumont, 2006; Zirkel, 2007). The RtI process is described as a tiered intervention approach that identifies students with learning disabilities dependent upon their response to evidence-based instruction (Ardoin, 2006; Daley et al., 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; VanDerHeyden, Witt, & Gilbertson, 2007). Essential to this process is that high-quality instruction or intervention is provided to meet a student's needs thus enabling educators the opportunity to make important educational decisions based on a student's learning rate over time and level of performance (Ardoin, Witt, Connell, & Koenig, 2005; Barnett et al., 2007; Daley et al., 2007; Fuchs et al., 2008; NASDSE, 2005; Shinn, 2007; VanDerHeyden et al., 2007). When a student does not respond to multiple tiers of increasing support, he or she is then considered eligible for special education services as a student with a learning disability (Ardoin, 2006; Ardoin et al., 2005; Barnett et al., 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Shinn, 2007).

In order to determine whether a student is making adequate progress and thus responding to an intervention, it has been important to measure a student's progress with a tool that is sensitive to instructional change over time (Cusumano, 2007). Curriculum-based measurement (CBM) is often used for this purpose and thus is often included in a school district's implementation of the RtI process in order to monitor the progress of the students across time (Ardoin, 2006; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Shapiro, 2004). Curriculum-based Measurement (CBM) is a fast, reliable, and valid measure of student progress (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994; Good & Jefferson, 1998; Hosp, Hosp, & Howell, 2007; Marston, 1989; Stecker, Fuchs, & Fuchs, 2005). Thus, CBM is a powerful tool that can be used within the process of RtI to monitor the progress of students and eventually help decide the necessity for special education services for individual students.

In addition to identifying students with special needs, CBM is also used to monitor students' progress towards meeting end of year academic expectations as determined through the use of state-mandated achievement tests (Crawford, Tindal, & Stieber, 2001; Helwig, Anderson, & Tindal, 2002; Hintze & Silberglitt, 2005; Jitendra et al., 2005; McGlinchey & Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Silberglitt et al., 2006; Stage & Jacobsen, 2001). The passage of the No Child Left Behind Act of 2001 (NCLB) placed increased emphasis on accountability through the use of achievement tests (Katsiyannis et al., 2007; McGlinchey & Hixson, 2004; Silberglitt et al., 2006; Silberglitt & Hintze, 2005). Specifically, NCLB mandated that a school's effectiveness would be determined based on that school's ability to obtain adequate yearly progress, as measured through statewide accountability testing (Katsiyannis et al., 2007; NCLB, 2001;

Silberglitt et al., 2006). Thus, many schools have turned to alternative forms of measurement to determine how to alter instruction of the appropriate skills in enough time to benefit students' performance on state-mandated achievement tests.

CBM is a sensitive measure of basic skill progress over time; moreover, it has been used as a dynamic indicator of overall academic growth and development (Hasbrouck, Woldbeck, Ihnot, & Parker, 1999; Hintze, Christ, & Keller, 2002; Hosp et al., 2007; Hintze & Silberglitt, 2005; Stecker et al., 2005). CBM is a logical tool for measuring progress towards academic goals, and extant research demonstrates high correlations between various CBM tools and various achievement tests which measure proficiency in these academic goals (Helwig et al., 2002; Hintze & Silberglitt, 2005; Jiban & Deno, 2007; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005). CBM can provide educators with information about a student's progress to meet end-of-year goals that are assessed on state-mandated tests (Helwig et al., 2002; Hintze & Silberglitt, 2005; Jiban & Deno, 2007; McGlinchy & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005). This information allows educators to change instruction in enough time to increase student knowledge and eventually performance on the state-mandated tests (Helwig et al., 2002; Jiban & Deno, 2007). When student performance improves, the accountability of the school also improves (Helwig et al., 2002; Hintze & Silberglitt, 2005; Jiban & Deno, 2007; Silberglitt et al., 2006; Silberglitt & Hintze, 2005).

Unfortunately, many educators remain unaware of the effectiveness of CBM as an indicator of student performance in the basic skills (Fuchs & Fuchs, 2001; Hosp & Hosp, 2003; Hosp et al., 2007). Because of unfamiliarity with CBM, teachers may be hesitant to use it within their classrooms in the manner recommended (Fuchs & Fuchs, 2001; Hosp & Hosp, 2003; Hosp et al., 2007). In a study conducted by Hasbrouck and colleagues (1999), a teacher reported concerns about using CBM within her special education classroom. The teacher reported that she regarded CBM as an additional chore during her day that would keep her from teaching, and she was concerned about the increased accountability that would be associated with her students' performance on these measures (Hasbrouck et al., 1999). If teachers are likely to have negative reactions to the introduction of CBM within their classroom, then they may be less likely to use the assessments at all, including using the information to monitor intervention implementation effectiveness.

The logic of acceptability as explained by Wolf (1978) states that if a person does not like a treatment, then he or she would be more likely to avoid the treatment. Years of research have indicated that the level of personal regard for a strategy will ultimately affect the degree to which a person uses it (Fuchs & Fuchs, 2001; Hosp & Hosp, 2003; Hosp et al., 2007). Therefore, it is important that teachers who are being asked to use curriculum-based measurement on a regular basis regard the strategy as appropriate, fair, and valid in order to ensure that the tool is being used as it was intended (Allinder & Oats, 1997; Eckert, Shapiro, & Lutz, 1995; Kazdin et al., 1981).

Overall, it is important that teachers regard CBM as an acceptable measurement system because it has met the needs present in the education system (Allinder & BeckBest, 1995; Fuchs, 2004; Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994; Hosp et al., 2007; Hosp & Hosp, 2003). CBM is used as an alternative method in assessing and diagnosing student ability, and to track student progress towards meeting end-of-year expectations (Foegen & Deno, 2001; Foegen, Jiban, & Deno, 2007; Francis et al., 2005; Fuchs & Fuchs, 2002; Fuchs & Fuchs, 2007; Hosp & Hosp, 2003; Jitendra et al., 2005; McCardle et al., 2008; Stecker et al., 2005). CBM has been found to have predictive validity when compared to many state administered year-end tests used for accountabilities purposes (Helwig et al., 2002; Hintze & Silberglitt, 2002; Jiban & Deno, 2007; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005). However, not all standardized assessments have been compared with CBM, basic skills assessments. In addition, the acceptability of these procedures has received little attention with teachers who have no prior experience administering the assessments.

The present investigation is a replication and extension of a previous study conducted by Connell (2006). In that study, the author developed a set of mathematics concepts and applications measures that correlated highly with a number of criterion assessments. In addition, the probes were rated highly by the classroom teachers who were provided with the results to benefit instruction. The present investigation replicated Connell (2006) by evaluating whether assessments of basic math computations skills and/or basic math applications skills were correlated with a different set of nationally-used, norm-referenced measures of student achievement. The earlier study was extended by using the validated probes as a measure of progress monitoring multiple times throughout the year. In addition, this investigation helped determine teacher perceptions of the tool when they themselves were left to administer and score the probes. Finally, this investigation explored whether teacher perceptions of the basic math skills assessments improved after its use in the classroom.

CHAPTER 2

LITERATURE REVIEW

Curriculum-based Measurement (CBM) is a fast, reliable, and valid measure of student progress (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005). However, the acceptability of these procedures has not been investigated with corrective education teachers who have had no prior experience with these measures. While CBM has been touted as a viable method for monitoring the progress of students towards meeting the end-of-year expectations (Helwig et al., 2002; Hintze & Silberglitt, 2002; Jiban & Deno, 2007; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005), its use has been limited within the education field as of yet. The reason behind its limited use within the education field has not yet been determined, despite the fact that CBM has been considered a cost-effective, time-efficient, valid and reliable method of monitoring a student's progress (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Hintze, Christ, & Keller, 2002; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005). In addition to all of these features, it is important that the measurement tool be viewed as acceptable by those who use it (Allinder & Oats, 1997; Eckert et al., 1995; Kazdin et al., 1981). With a positive view of CBM, individuals within the education field may be more likely to use CBM and use it effectively (Allinder & BeckBest, 1995; Fuchs, 2004; Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994; Hosp et al., 2007; Hosp & Hosp, 2003). This study evaluates the predictive and concurrent validity of math applications and math computation curriculum-based assessments with student performance on a norm-referenced, state administered assessment. This study also evaluates whether exposure to and use of CBM practices improves teacher acceptability of those practices.

Curriculum Based Measurement

History of CBM

Curriculum based measurement was originally developed to test the effectiveness of a special education intervention model using a data-based program (Deno, 2003; Hosp et al., 2007). Prior to the development of CBM by Deno, Marston, Robinson, Shinn, Tindal, Wesson, and Fuchs, students' progress was monitored by measuring the students ability to master each objective taught in class (Fuchs, 2004). Learning was considered to be a series of short-term accomplishments that would eventually develop into "broad competence" (Fuchs, 2004, p. 188). In response to this perception, Deno and his doctoral students (1985) helped to develop this form of curriculum based assessment in a response to the "apparent limitations of commercially distributed achievements tests" (Deno, 1985, p. 219). He noted that measuring student achievement is essential to evaluate the success of an educational program. In addition, he added that while these limited commercially distributed tests were used for making instructional decisions within schools, it was yet unclear that these tests were viable for this role (Deno, 1985). Deno noted that, more commonly, teachers were found to use observation of students, instead of results from standardized tests, when assessing students' progress. However, Deno also indicated that teachers were found to be "biased toward judging that more students had successfully attained objectives than [actually] had" (p. 220). Deno responded to these difficulties by working to develop the initial forms of curriculum-based measurement. These efforts made student achievement data more essential and efficient for use in daily teacher decision making. The development of CBM allowed for the monitoring of a student's progress by developing measures that integrated various skills that are essential to the annual curriculum that was used each week (Fuchs, 2004). Because each weekly test measured the same concepts and was at the same level of difficulty, the slope of a student's progress on each weekly assessment could be used to define the rate of learning (Fuchs, 2004).

Deno's (1985) initial efforts were in the subject area of reading. Deno considered the goal of reading to be the comprehension of text. Thus, he intended to determine a means of measuring the comprehension of text in a way that fit his four point criteria. The assessment had to be reliable and valid, simple and efficient, easily understood, and inexpensive (Deno, 1985). However, Deno did not think that asking questions about a reading passage would be reliable, valid, simple, and efficient method of assessment. He then decided to use three initial tasks: a "read aloud" task, a brief comprehension task, and a word meaning task (Deno, 1985). Students were asked to read aloud from isolated word lists or from text passages, supply words deleted from text (cloze procedure), or say the meanings of words underlined in text in order to assess their reading progress. Deno noted that each of these tasks (except for the word meaning task) was found to be highly correlated with performance on the standardized, norm-referenced tests, ranging from .70 to .95. In addition, Deno noted that the task of reading aloud from text was able to consistently identify students who were placed in special education resource programs.

Overall, Deno (1985) noted a number of advantages of CBM for use in the school setting. It provided important information in a clear, simple, and meaningful way that it could be used to make educational decisions (Deno, 1985; Fuchs et al., 1994; Hosp & Hosp, 2003; Hosp et al., 2007). It was also sensitive to growth in a student's progress over short periods of time thus enabling a teacher to evaluate his or her efforts and effectiveness (Cusumano, 2007; Deno, 1985; Deno, 2003; Fuchs et al., 1994; Hosp & Hosp, 2003; Hosp et al., 2007; Stecker et al., 2005). As such, a teacher could make changes quickly in response to identified problem areas. Additionally, because of the relative ease of administration, it was possible for CBM measures to be administered frequently, improving the quantity of data available to teachers for decision making (Cusumano, 2007; Fuchs et al., 1994; Deno, 1985; Deno, 2003; Hosp & Hosp, 2003; Hosp et al., 2007; Stecker et al., 2005). Because the cost of interventions was important for schools to consider, Deno (1985) further recommended the use of CBM as a cost-effective strategy that did not require the purchase of additional materials and did not require a great deal of time to administer, as compared to commercially distributed achievement tests. Therefore, CBM was touted as "the promise of the future" in assessment practices (Deno, 1985, p. 231).

As a result of Deno's (1985) initial research in this area, additional studies have been conducted to continue the thorough exploration of this assessment strategy within the classroom. The results of this exploration have identified CBM as a reliable and valid measure of a student's academic progress towards meeting established criteria, such as curriculum goals using standardized observational procedures (Deno, 2003; Fewster & MacMillan, 2002; Fuchs et al., 1994; Fuchs et al., 1991; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005; Thurber, Shinn, & Smolkowski, 2002; VanDerHayden, Witt, Naquin, & Noell, 2001).

Technical Adequacy of CBM

Deno's (1985) work created a new form of assessment that could easily be used within the classroom to monitor the progress of students. However, in order for the assessment to be considered a curriculum-based measure, there were a number of characteristics that needed to be present (Deno, 1985; Marston, 1989). The criteria established were that the measure had to be: (1) tied to the curricula, (2) of short duration to facilitate frequent administration by staff, (3) capable of having many multiple forms, (4) inexpensive to produce, and (5) sensitive to instruction (Marston, 1989, p. 30). It was noted that CBM was a method of direct measurement, or assessing a behavior by observing its occurrence (Marston, 1989).

CBM procedures underwent a great deal of investigation and were found to have high levels of technical adequacy (Fewster & MacMillan, 2002; Good & Jefferson, 1998; Hosp et al., 2007; Marston, 1989). Analyses of the extant literature conducted by Marston (1989) and Good and Jefferson (1998) revealed numerous studies demonstrating both reliable and valid CBM procedures in assessing reading, spelling, written expression, and math.

It was noted that reading stories aloud (or oral reading fluency) was consistently found to be a valid measure of student achievement in reading when correlated with various standardized, norm-referenced, criterion-based, standardized assessments measuring reading mastery such as the Stanford Diagnostic Reading Test, the Woodcock Reading Mastery Test, and the Reading Comprehension subtest of the Peabody Individual Achievement Test (Marston, 1989). Marston (1989) noted correlation coefficients ranging from .63 to .90 with most coefficients being above .80. Reliability coefficients for these measures were determined through test-retest reliability, parallel form estimates, and interrater agreement; the coefficients ranged from .82 to .97 with most coefficients above .90 (Marston, 1989).

In the assessment of spelling, Marston (1989) noted that most studies scored performance using either the number of words spelled correctly or the number of correct letter sequences and compared that to various criterion measures, such as the Stanford Achievement Test, the Test of Written Spelling, and the Peabody Individual Achievement Test. Both methods of scoring were repeatedly found to be valid measures with correlations ranging from .81 to .95 (Marston, 1989). The same measures of reliability were utilized, and it was determined that the estimates ranged from .72 to .97 for both forms of scoring (Marston, 1989).

Written expression CBMs were also investigated and compared to the Test of Written Language, the Developmental Sentence Scoring System, and the Language subtest of the Stanford Achievement Test (Marston, 1989). Written expression CBM studies investigated multiple scoring methods: total words written, words spelled correctly, correct letter sequences, and mature words (Marston, 1989). All of these methods were highly related to the criterion measures with total words written having correlations ranging from .41 to .84 (Marston, 1989). Reliability coefficients from the multiple methods were mostly above .70, ranging from .42 to .91 for total words written (Marston, 1989).

The assessment of math CBM by Marston (1989) revealed more limited technical adequacy than the other subject areas. Math CBM was found to have lower relations with the criterion measures of the Metropolitan Achievement Test (MAT) Problem-Solving and the MAT Math Operations with few correlations exceeding .60 (Marston, 1989). To explain this, Marston (1989) noted that there had already existed a general concern about the appropriateness of published math tests because of limited content. However, Marston (1989) indicated that single administrations of CBM math probes have demonstrated reliability using test-retest, parallel form estimates, and interrater agreement.

Because of these favorable reports of the technical adequacy of various CBM methods, Marston (1989) recommended their use in various areas of education. In addition, research has continued to validate the positive attributes of curriculum-based measurement. Fuchs and Fuchs (2002) reviewed research on reading CBM that demonstrated its psychometric tenability and its use within instructional planning, evaluation, and identification procedures. They noted that there has been evidence of adequate predictive validity and construct validity when compared to commercial reading tests, and to teachers' judgments of reading competence (Fuchs & Fuchs, 2002). There has been evidence of the ability of CBM to discriminate students based on their scores on reading CBM with respect to special education status and grade (Fuchs & Fuchs, 2002). In addition, there has been evidence of stability and interrater agreement (Fuchs & Fuchs, 2002).

Good and Jefferson (1998) extended the work of Marston by being more specific about the criterion-based analyses studies included in the analysis of curriculum-based measurement. For example, Good and Jefferson only included studies that compared scores within the same grade level and those studies that used the most frequently recommended CBM scoring rules. The authors noted that the most information was available about CBM reading. The validity coefficients in CBM reading ranged from .60 to .80, thus indicating that CBM reading had construct validity. Fewer studies were available that explored CBM for written expression. However, the authors concluded that modest support for construct validity in CBM written expression was obtained in the elementary grades, ranging from .60 to .80. When exploring the validity of CBM math, the authors were only able to use information from the third and fifth grades. Only modest support was obtained for the construct validity of CBM math.

Through the use of various measures to determine both the validity and values of CBM, Good and Jefferson (1998) evaluated the science of CBM as well as the ethics of the measurement process. They determined that when CBM is used within a problem-solving process, there is a valid basis for both test interpretation and use with students.

A meta-analysis conducted by Foegen, Jiban, and Deno (2007) which focused on math CBM investigated the two common concerns for math CBM identified by Fuchs (2004). They were interested in investigating the strategies used to develop CBM measures, and the stages of research that could establish the practicality of CBM measures (Foegen et al., 2007). This information could then be used to identify what areas of mathematics progress monitoring were in need of further research to ensure that the best possible forms of progress monitoring were being used to assess student achievement. Two broad approaches have been used in developing CBM measures, particularly in the area of mathematics: curriculum sampling and robust indicators, as identified by Foegen and colleagues (2007). In curriculum sampling, the CBM measures are developed by creating samples that are representative of the year's mathematics curriculum. This method has been used for computation probes as well as concepts and applications probes. The authors note that while there is a direct link to the curriculum with this approach and this may help with the design of effective remediation strategies, it can only be used across one year and thus cannot show a student's growth across multiple years (Foegen et al., 2007). It will, therefore, be necessary to develop multiple CBM systems with each linked to the specific curriculum being used at that time. The second approach, which the authors preferred, is the robust indicators approach. This is when measures are used that broadly define proficiency in mathematics, and thus may not be representative of a particular curriculum but still correlate with various criterion measures (Foegen et al., 2007). This method is preferred because it can be used across multiple grade levels, can model a student's growth over multiple years of learning, and identifies aspects of core competence in mathematics. It is, however, less useful in providing diagnostic information concerning the strengths and weaknesses of a student in particular areas.

The thirty-two studies reviewed by Foegen and colleagues (2007) were divided into three categories. The authors used a three-stage continuum of CBM research that was initially described by Fuchs (2004). Stage 1 research was defined as a study that explored the technical adequacy of CBM. More specifically, it emphasized the reliability and criterion validity of the measures (Foegen et al., 2007; Fuchs, 2004). Stage 2 research examined the slopes that were generated by CBM used for continuous monitoring (Foegen et al., 2007; Fuchs, 2004). Stage 3 research examined teachers' use of CBM when developing instruction meant to improve student achievement (Foegen et al., 2007; Fuchs, 2004). Studies that fell under all three categories were included in the author's review of the literature.

Foegen and colleagues (2007) indicated that the majority of studies that have been completed have focused on elementary mathematics, as opposed to early mathematics or secondary mathematics. Within these elementary school studies, there was no clear pattern concerning the method used to develop the probes, while in early mathematics studies, it was clear that the robust indicators method was the only method utilized to develop CBM. However, studies that investigated the difference between the robust indicators method and the curriculum sampling method indicated that both methods were able to discriminate between students of different mathematics ability levels.

An important finding by Foegen and colleagues (2007) was that “the strength of the correlations for a CBM reading measure exceeded those for either CBM mathematics measure [i.e., computation or concepts and applications] on four of the five criterion measures” (p. 136). The authors also noted that the majority of studies that have been completed are categorized as Stage 1 research. The two studies that the authors reviewed that were considered Stage 3 research did result in increases in student achievement, but more positive results were obtained when CBM was paired with additional support, such as consultation or self-monitoring.

Fuchs and Fuchs (2002) also noted that CBM has been found to be a good instrument for monitoring individual change over time. This was established because CBM provided equal scaling of individuals on the behavior measured over time, the construct and difficulty level stayed constant across time, and there were a sufficient number of alternate forms available. Because of these characteristics, the authors concluded that CBM could be used to model academic growth and develop a trajectory reflecting a student’s change progress.

Overall, Fuchs and Fuchs (2002) recommended that CBM be used to enhance teacher planning and ultimately student learning. They indicated that CBM could help teachers set ambitious goals for the students, determine when adaptations were necessary for more growth, and provide ideas for teaching adjustments.

The research demonstrates that CBM measures have a high degree of technical adequacy (Foegen et al., 2007; Fuchs & Fuchs, 2002; Good & Jefferson, 1998; Marston, 1989). CBM has proven to be a useful and, arguably, an essential tool for the development and monitoring of an effective educational program for students (Foegen et al., 2007; Fuchs & Fuchs, 2002; Good & Jefferson, 1998; Marston, 1989).

CBM and the Individuals with Disabilities Education Improvement Act of 2004

The identification of students' eligibility for special education is a common practice within the education system. School psychologists have reported that they spend the majority of their professional time assessing students to determine eligibility for special education (Agresta, 2004; Eckert, Shapiro, & Lutz, 1995; Lichtenstein & Fischetti, 1998; Peterson & Shinn, 2002). The prior federal definition of a learning disability stipulated that a severe discrepancy be evident between a student's achievement, and the student's intellectual ability (Francis et al., 2005; Fuchs & Fuchs, 2007; Fuchs, Fuchs, & Zumeta, 2008; Hilton, 2007; McCardle et al., 2008; Peterson & Shinn, 2002). An observed discrepancy between the two scores obtained has been interpreted by the education system as the means for measuring a student's unexpected underachievement; however, there has been waning support for the reliability and validity of this practice (Francis et al., 2005; Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klinger, 2007; McCardle et al., 2008; Willis & Dumont, 2006). It has become clear that when using this process, students are often identified in grades three through five (Fuchs & Fuchs, 2007; Fuchs et al., 2008; McCardle et al., 2008). Therefore, many of these students fall behind in the general education curriculum before they meet the criteria for classification as a student with a learning disability. Thus, many people have identified this discrepancy model as a "wait-to-fail" model (Fuchs & Fuchs, 2007; Fuchs et al., 2008; McCardle et al., 2008). "In other words, children have to fail (repeatedly) in general education before they are recognized as having a disability and given access to more appropriate instruction in special education" (Fuchs et al., 2008, p. 116).

Another major concern of the discrepancy approach is the estimation of the discrepancy (Francis et al., 2005; Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klinger, 2007; McCardle et al., 2008; Willis & Dumont, 2006). Additionally, Hosp and Hosp (2003) have indicated concerns that many norm-referenced tests lack any degree of treatment validity. The lack of treatment validity questions whether the results of an IQ or achievement test could help guide instruction and thus improve student performance. Finally, minority students are historically over represented in special education settings causing critics to question the discrepancy model in that process (Harry & Klingner, 2007).

The most recent federal legislation, the reauthorization of Individuals With Disabilities Education Improvement Act (IDEIA 2004; P.L. 108-446), has allowed states and districts to consider the use of an alternative method for identifying students with learning disabilities (Fuchs & Fuchs, 2007; IDEIA, 2004; Zirkel, 2007). Another possible method for determining eligibility for special education services permitted in IDEIA has been labeled the response to intervention (RTI) approach, and it prescribes that special education eligibility be based upon a student's lack of response to high quality, evidence-based instruction (Ardoin, 2006; Barnett et al., 2007; Daly et al., 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Harry & Klingner, 2007; Hilton, 2007; Shinn, 2007; Willis & Dumont, 2006; Zirkel, 2007). Within this model, if a student is exposed to high-quality, evidence-based intervention and if through a process of providing a continuum of increasingly intensive supports that student has not responded to the interventions, then he or she should be considered eligible for a special education eligibility determination

(Ardoin, 2006; Daley et al., 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; VanDerHeyden, Witt, & Gilbertson, 2007).

The tiered system (i.e., increasing levels of support) of RtI dictates the level of intensity of instruction that is provided to students based on their individual needs (Fuchs et al., 2007; Fuchs et al., 2008; NASDSE, 2005; Spear-Swerling, 2008). The first tier, which consists of approximately 80% of the school population, receives the core instructional interventions (Fuchs et al., 2007; Fuchs et al., 2008; NASDSE, 2005). Each student is exposed to a scientifically validated curriculum, and interventions used at this level are oriented toward whole-group instruction (Fuchs et al., 2007; Fuchs et al., 2008; NASDSE, 2005; Spear-Swerling, 2008). Those students who continue to struggle within the content areas may be identified for additional supports in Tier 2. In the second tier, which is composed of approximately 15% of students, supplemental instruction is provided to the students identified as struggling in Tier 1 (Fuchs et al., 2007; NASDSE, 2005; Spear-Swerling, 2008). Systematic, evidence-based interventions are used with students in this second tier to assist in the improvement of academic skills (Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klinger, 2007; NASDSE, 2005; Spear-Swerling, 2008). The progress of these students is to be monitored so that instruction can be altered based on the response to the intervention (Fuchs et al., 2008; Harry & Klingner, 2007; Hilton, 2007; NASDSE, 2005). If students continue to struggle, they then may be moved to the third tier, consisting of approximately 5% of the student population (Fuchs et al., 2007; NASDSE, 2005; Spear-Swerling, 2008). In this tier, a student receives intensive instructional interventions to increase the rate of progress (Fuchs et al., 2007; Fuchs et al., 2008; Harry & Klinger, 2007; NASDSE, 2005; Spear-Swerling, 2008). The services in this tier might include special education programs or other long-term interventions (Fuchs et al., 2007; Fuchs et al., 2008; NASDSE, 2005; Spear-Swerling, 2008).

The National Association of State Directors of Special Education (NASDSE) recommends that, within an RtI framework, entire school populations be screened in order to identify students struggling more than others (NASDSE, 2005). In addition, the NASDSE notes that ongoing assessment of proficiency is essential to the success of this system. The data obtained from progress monitoring tools can inform instruction at each tier and identify the appropriate level of service needed for each student. Thus, academic progress must be monitored frequently within the RtI model in order to ensure the interventions are sufficient to help the student maintain the same level as his or her peers (Fuchs et al., 2008; Harry & Klingner, 2007; Hilton, 2007; NASDSE, 2005).

While some consider the response to intervention approach and the discrepancy approach to be diametrically opposed, others propose that both methods should be used together to meet the diagnostic demands of IDEIA 2004 (Fiorello et al., 2006; Hale et al., 2006; Naglieri & Kaufman, 2008; Willis & Dumont, 2006). The law does not clearly determine which method should be used in the identification process and leaves the state or district with the responsibility of deciding which would be the most appropriate to fit their needs (Fuchs & Fuchs, 2007; IDEIA, 2004; Zirkel, 2007). However, the law does clearly state that a variety of valid and reliable assessment tools be utilized to gather information about the student (Naglieri & Kaufman, 2008, Zirkel, 2007). CBM has been used as a measurement tool that supports the use of RtI within the identification process (Ardoin, 2006; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Shapiro, 2004). The monitoring of a student’s progress towards meeting the end-of-year goals is an essential piece of the RtI process within the schools (Fuchs et al., 2008; Harry & Klingner, 2007; Hilton, 2007; NASDSE, 2005). Important decisions are made based on a student's learning rate, or growth in achievement over time compared to prior performance and peer rates, and level of performance, or a student's relative standing on some dimension of achievement compared to a criterion- or norm-referenced expected performance (Ardoin et al., 2005; Barnett et al., 2007; Daley et al., 2007; Fuchs et al., 2008; NASDSE, 2005; Shinn, 2007; VanDerHeyden et al., 2007). A student is considered for special education when he or she "not only performs below the level demonstrated by classroom peers but also demonstrates a learning rate substantially below that of classmates" (Fuchs & Fuchs, 2002, p. 77). Curriculum-based measurement is a progress monitoring tool that is standardized with documented reliability and validity, and fits the needs for progress monitoring within this approach (Fuchs et al., 2008). Because CBM is developed based on goals for the year's end, they are useful measures that enable educators to monitor a student's progress throughout the year to determine both a student's learning rate and level of performance (Ardoin, 2006; Ardoin et al., 2005; Cusumano, 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Shapiro, 2004).

In order to further explore the use of CBM specifically as a progress monitoring tool within an RtI three-tiered framework, Fuchs and colleagues (2007) explored math screening tools as a means of predicting those students who would be later identified as having a math disability. In their study, first grade students were administered various multiple skill math screening measures in September. The students were then administered weekly CBM probes to further monitor their progress. The findings of this study indicate that multiple skill screeners provided a strong correlation to outcomes of mathematics assessment. In addition, this study concluded that CBM computation probes may be useful in obtaining valid information about math competence development across first grade. Thus, math CBM may be helpful in determining whether a student is responding to intervention within the second and third tiers of the framework.

School Extensions of CBM to Evidence-Based Practices

CBM has found widespread use throughout the education system because of its confirmed reliability and validity properties with various school aged populations (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Hosp, et al.; Marston, 1989). Studies have shown that with objective ongoing assessment information such as CBM, teachers are given the opportunity to determine when a student's progress is inadequate and adjust instructional programs more frequently to meet the needs of the student or needs of the whole class (Cusumano, 2007; Deno, 1985, 2003; Fewster & MacMillan, 2002; Fuchs, 1992; Fuchs et al., 1991; Hosp et al., 2007). The hope is that these adjustments may ultimately enable students to achieve better outcomes (Cusumano, 2007; Deno, 1985, 2003; Fuchs, 1992; Fuchs et al., 1994; Hosp et al., 2007). Thus, CBM has been researched in coordination with other established best practices in order to determine whether the addition of progress monitoring would strengthen other procedures (Allinder & BeckBest, 1995; Calhoon & Fuchs, 2003; Fuchs et al., 1991; Fuchs et al., 1994; Fuchs & Fuchs, 2001).

Teacher Consultation

Consultation has been defined by Caplan as the process of interaction between two professional persons working to plan or implement a program in order to assist clients (Erchul & Martens, 2002). Fuchs, Fuchs, Hamlett, and Stecker (1991) found that CBM, when used with

Expert Systems (ExS), a computer program of instructional consultation, increased student achievement. The 33 teacher participants were separated into one of three groups: CBM with the ExS recommendations, CBM without ExS recommendations, or the control group that did not use CBM. Each teacher selected two students who were chronically low achieving in mathematics, in second through eighth grade to monitor progress. Students were administered a 25 problem CBM test twice weekly for 20 weeks. Then, depending upon their assignment to the various treatment groups, the teacher used the ExS computer program. To use the computer program, teachers entered information about the pattern of the student's graphed outcome data, the mastery levels, the percentage of correct digits for the problem type, the nature of the previous instruction, judgments about the quality of the student's daily work, and curricular priorities. From this, the computer program would recommend an instructional adjustment and detailed instructions for how to implement the adjustment.

The results of the study revealed that teachers who used CBM had greater results on student achievement than teachers who did not use either CBM or consultation, and thus did not monitor student progress (Fuchs et al., 1991). The greatest improvement in student achievement, though, was found in the group that utilized both CBM and school-based consultation. The authors determined that objective ongoing assessment, such as CBM, may be an important tool in encouraging teachers to make appropriate instructional adjustments throughout the academic year. However, this study determined that without consultation, CBM may not be enough to encourage the best possible instructional adjustments. The authors speculated that with consultation, teachers have a better idea of what to adjust and how to modify instruction rather than relying on their knowledge to interpret the results of CBM data and use that information to positively affect instruction.

Another study conducted by Fuchs, Fuchs, Hamlett, Phillips, and Bentz (1994) further investigated whether providing recommendations to teachers along with results of math CBM would have an impact on student achievement. Forty general education teachers in a southeastern, urban school district with at least one student in their second through fifth grade classrooms with an identified learning disability in math participated in this study. They were randomly assigned to one of two conditions, each involving the use of CBM to monitor the progress of students. Each teacher administered a weekly CBM for 25 weeks. The software used provided the teachers with a graph of the student's progress as well as a skills profile demonstrating mastery status on each type of problem included in the year's curriculum. Twice a month, the teachers were provided with a computer-generated copy of the student's graph and skills profile and a summary of the performance of the class. Teachers in the first condition were provided with a report that only described the performance of the students without any recommendations, while teachers in the second group were provided with descriptions as well as instructional recommendations.

Results of this study by Fuchs and colleagues (1994) indicated that when CBM decision-making strategies have a class-wide focus, as in this study when the teachers were provided with descriptions about class-wide performance, there is an opportunity to improve student achievement. The results of this study also demonstrated that when general educators were provided with CBM information, along with instructional recommendations, students achieved at a greater rate than those students who had not been administered CBM. These students achieved at a greater rate regardless of their classification as learning disabled, low achieving, or average achieving. The effect size for those students in the instructional group versus contrast was .43 while the effect size for those students in the non-instructional group versus contrast was .16. However, the authors did indicate that it may be necessary to consider additional forms of adaptations dependent upon the needs of the individual students.

A study by Allinder and BeckBest (1995) also investigated the effects of consultation with special education teachers who were implementing a data-based instructional system. The system required staff monitoring of math computation skills using the Math Computation Test-Revised (MCT-R). Eighteen teachers monitored the math computation progress of two students in their classrooms. Eight of these teachers received bimonthly consultation that was university-based. The remaining ten participants self-monitored their use of CBM. The teachers were trained in the use of this method and were told to administer a CBM on a biweekly basis, but the teachers did vary on how often students actually took the tests. The information was then graphed and teachers were to review the graph twice monthly and decide whether the student was making adequate progress towards the year-end goal and should continue the instructional program. The teacher could also decide if the student was making better than expected progress and then raise the goal, or whether the student was making less than expected progress and then modify the instructional program. Every other week the teachers would review the procedures that were completed in one of two ways. Either the teacher would meet with their university-based consultant who would answer any questions the teacher may have, or the teacher would complete a questionnaire that consisted of questions regarding the critical components of CBM.

The results of this study by Allinder and BeckBest (1995) indicated that achievement was made by the students taught by teachers in each of the groups, consultation and self-monitoring. The gains made by these students were not significantly different from each other. The authors indicated that the small number of participants did, however, limit the effect size of the variables; thus it is possible that with a larger sample size different results may be obtained. Overall, it does appear that student progress was made, although it is not clear that it was a direct result of CBM, university-based consultation, or self-monitoring of the teachers.

The results of aforementioned studies seem to indicate that while CBM is an effective tool for monitoring a student's progress, it can become a more useful tool that can even potentially improve the effect of such research-based practices as consultation (Allinder & BeckBest, 1995; Fuchs et al., 1991; Fuchs et al., 1994). Consultative recommendations may be helpful even in the form of computer distributed recommendations matching student responses (Fuchs et al., 1991; Fuchs et al., 1994). However, without the administration of these progress monitoring tools, it is not possible to learn how to better effect student achievement (Fuchs et al., 1991; Fuchs et al., 1994).

Peer Tutoring

Fuchs and Fuchs (2001) noted the importance of CBM as a classwide measurement tool. They reported that CBM provides helpful information in a way that is able to focus teacher attention on individual students who were in the greatest need. In developing a peer tutoring program (Math Peer-Assisted Learning Strategies; Math PALS), Fuchs and Fuchs (2001) included weekly and biweekly monitoring in the form of CBM as a means for the classroom teacher to understand better the progress students made towards meeting the end-of-year goal. The data obtained through the use of CBM enabled Math PALS to be designated as an "effective practice" with validity and reliability by the Program Effectiveness Panel in the U.S. Department of Education (Fuchs & Fuchs, 2001). The authors noted that one of the most important aspects of integrating this research-based practice into a classroom is strongly focusing on student outcomes (Fuchs & Fuchs, 2001). The use of CBM allows for a continued focus on the goal and thus allows for effective monitoring of student progress (Cusumano, 2007; Deno, 1985, 2003; Fewster & MacMillan, 2002; Fuchs, 1992; Fuchs et al., 1991; Hosp et al., 2007). In addition to being a central piece of the intervention, CBM allowed for the determination of Math PALS as an effective classwide intervention strategy (Fuchs & Fuchs, 2001).

A later study by Calhoon and Fuchs (2003) further investigated the use of Math PALS with three high school math teachers who taught 92 students in ninth through twelfth grades identified as having a learning disability in mathematics. The classrooms of these teachers were evenly divided into a treatment group: those who participated in a PALS/CBM program within math class, and those in a control group who utilized a math workbook. All of these students were administered the Math Operations Test-Revised (MOT-R), the Math Concepts and Applications Test (MCAT), and the Tennessee Comprehensive Achievement Test (TCAP) both before and after the integration of this program into the curriculum in order to assess the math achievement of the students in the study. The students in the treatment group completed a questionnaire asking their opinions about both PALS and CBM. The teachers in the treatment group also completed a similar questionnaire. The teachers trained the students on four separate occasions on the appropriate usage of the PALS program. The computation portion of PALS occurred the first eight weeks of the study, while the concepts/applications portion of PALS occurred for the next eight weeks. The program took place two days a week and lasted 30 minutes per session. A CBM probe was administered one time a week in order to track the students' progress.

The results of the study revealed the PALS/CBM group outperformed the control group on the assessment of computation skills, while both groups significantly improved in the area of concepts/applications and on the TCAP. Based on the answers to the questionnaires, it was revealed that both the teachers and the students reported positive attitudes about the use of PALS and CBM in their class. The authors concluded that peer tutoring could be a useful intervention for secondary students with disabilities. However, it was necessary to use CBM to effectively determine the progress of the students as a result of this intervention.

Current Use of CBM within Schools

Curriculum-based measurement has been identified as a highly reliable and valid measure of student progress in the many different subject areas (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Good & Jefferson, 1998; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005). These procedures have been identified as useful in many different areas in contemporary classrooms (Fuchs, 2004). However, there is very little evidence that these procedures are currently being used by teachers and school psychologists (Hasbrouck et al., 1999).

One study that investigated the use of CBM within the classroom was completed by Allinder (1995) who examined the relationship between the efficacy of a teacher and the use of CBM. This study determined whether a teacher's personal and teaching efficacy would affect his or her use of CBM and how it might affect the academic growth of the students in the class. The author hypothesized that because teachers with high efficacy are more likely to implement new educational practices, these teachers might be more likely to use CBM as a regular practice. Nineteen special education teachers participated in this study, each choosing two students with whom they used CBM. Each of the students had a current Individualized Education Program (IEP) in the area of math. The students were administered the Math Computation Test-Revised both at the beginning and the end of the study to assess their achievement. The teachers were administered the Teacher Efficacy Scale at the beginning of the study to assess both teaching and personal efficacy. CBM in math were administered to the students twice a week, and the teachers reviewed the graphed performance of the students two times a month. Approximately eight weeks into the study, the Math-Modified Accuracy of Implementation Rating Scale (M-MAIRS) was completed by the project staff in order to determine the adequacy of CBM implementation by the teachers.

Results of this study by Allinder (1995) revealed that there was an effect of personal and teaching efficacy on both the implementation of CBM and student achievement. Teachers with high personal efficacy and high teaching efficacy did increase the end-of-year goals for the students more often than those teachers with low degrees of efficacy, with coefficients of determination at .26 and .42 respectively. Teachers with high teaching efficacy also set goals that were more ambitious than teachers with low efficacy with a coefficient of determination at .24. Students of teachers with high personal efficacy achieved better in the number of digits correct, the number of problems correct, and the slope of progress, than students with teachers having low personal efficacy. However, no such results were obtained when comparing teachers with high teaching efficacy versus low teaching efficacy. Thus, the authors concluded that high personal and teaching efficacy will have a positive influence on the students in the classroom. In addition, higher efficacy may result in more appropriate use of CBM within the classroom which may positively affect the performance of the students.

Hasbrouck, Woldbeck, Ihnot, and Parker (1999) related the experiences of one special education teacher who was new to the use of CBM in the classroom. She related that she was angry at the notion of a change within her typical routine and opposed to this practice in the beginning. She was being forced by her school district to integrate these procedures into the curriculum. She noted that she was sure she would not have enough time to adequately teach her curriculum, and that CBM would take time away for regular assessment. She added that she was fearful about the increased accountability that would be associated with having data as proof of a student's progress, or lack thereof. However, she began to use CBM and found the procedures to be very helpful.

This teacher noted three particular reasons why her skepticism changed to implementation (Hasbrouck et al., 1999). First, she related that graphing the results of CBM provided a powerful communication tool of data-based evidence for the parents to understand the progress of the students. Second, this teacher noted that when a student is making negative progress, she was able to identify this trend immediately and carefully observe if it continued. The teacher could begin to prepare instructional changes in the event that the negative progress continues. This also allowed for the teacher to use instructional time more efficiently by spending more time with those students making less progress. Finally, the teacher noted that she could easily observe if an instructional change had not been beneficial to the progress of the student, and she was able to make any necessary changes. The teacher added that she was able to use CBM results effectively because of the support she received from other teachers and the school's administration. In order to assist the teachers with the usage of this program, there was a district wide training and monitoring program put into place.

Overall, from this one teacher's experience, it appeared that a CBM program could be properly integrated into and appreciated in the classroom of a teacher with no prior experience (Hasbrouck et al., 1999). However, it must be noted that proper support was necessary to ensure a more positive experience for the teacher.

Many times, the support that is provided to teachers in integrating a CBM program comes from the school psychologist. Shapiro, Angello, and Eckert (2004) used a national survey of school psychologists to determine school psychologists' knowledge, use, and attitudes toward curriculum-based assessments (CBA) in everyday practice. A similar survey had been completed ten years prior to this study, and the authors compared the current results to the earlier results.

A total of 517 school psychologists participated in this Shapiro and colleagues' (2004) study by completing a 14-item questionnaire that had been mailed to them or they completed the survey online. The questionnaire obtained demographic information about the participants, asked specifically about their usage of CBA, explored their knowledge of CBA practices, and probed their attitudes towards the procedures.

The results of the study by Shapiro and colleagues (2004) revealed an increase in the percentage of people using CBA in practice, from 46% to 54%. Of that 54%, there were 32% that indicated a frequent level of usage. The highest percentage of non-users in both the 1989-1990 sample and the 1999-2000 sample reported that they had insufficient training in the technique. There was a significant increase in the percentage of responders who indicated that they had exposure to CBA in their graduate training, from 18% to 57%. There was no significant difference in the reported acceptability and attitudes about the use of CBA. Thus, there was no reported change in the attitude of school psychologists towards using CBA. Overall, the authors suggest that CBA may be becoming more prevalent in contemporary school psychology practice. However, it is possible that there is room for further use and greater approval of these methods within the classroom.

Teacher Acceptability of CBM

An intervention is only beneficial when it is put into practice. However, there is never a guarantee that a teacher will choose to use an intervention, even when it appears to be necessary and useful for student achievement. Wolf (1978) noted that if a participant in a study does not like a treatment, it is likely he or she may avoid the treatment. The treatment or intervention will not be able to impart its benefit on the participant. Wolf indicated that in order for a treatment to be used effectively it must have some form of social validity, which he defined as social significance of a goal, social appropriateness of a procedure, and social importance of an effect. He noted that despite the subjective nature of social acceptability, it should still be recognized as an important part of intervention. He indicated that the acceptability of an intervention or treatment will be related to the effectiveness of this intervention or treatment, as well as the probability that the program will be used and supported by the participants.

Kazdin (1977) indicated that often in clinical research small changes in behavior may not be the best possible criterion for treatment success. He noted that behaviors often have a level of social acceptability and therefore the presence or absence of a behavior may not be the best estimate of a treatment's success. This concept can also be applied to a school setting. Kazdin indicated that the concept of social validation should play an important role in determining the effectiveness of treatment. He defined social validation broadly as the social acceptability of a program. In a later study, Kazdin, French, and Sherick (1981) defined social validity as "judgments by laypersons, clients, and others of whether treatment procedures are appropriate, fair, and reasonable for the problem or client" (p. 493). Kazdin (1977) added that this concept can help bring together experimentation with practice. That is, recognizing that the social importance of a treatment may or may not make this treatment more likely to be used in a practical setting is important for experimenters to understand (Kazdin, 1977).

Just as social validity or acceptability is important within a clinical setting, it is important within a classroom setting (Allinder & Oats, 1997; Chafouleas, Riley-Tillman, & Eckert, 2003). Various studies have been conducted determining whether a procedure, such as CBM, can be considered acceptable by teachers or other professionals (Allinder & Oats, 1997; Chafouleas et al., 2003; Eckert, Hintze, & Shapiro, 1997; Shapiro & Eckert, 1994). This has often been described as treatment acceptability in which the variable of interest has been participants' evaluations of treatments through the use of various rating scales (Eckert, Hintze, & Shapiro, 1999). However, no study has yet looked at the acceptability of CBM both before and after its initial use, as rated by corrective education teachers actively involved in the integration of the procedure.

Eckert, Shapiro, and Lutz (1995) determined the perspectives of general and special education teachers towards different types of assessments. Participants were asked to read a case study and that presented data obtained either using curriculum-based assessment (CBA) or published norm-referenced tests (PNRT). It was determined that these teachers consistently rated CBA as a more acceptable method of assessment. These authors concluded that teachers who found an assessment process to be more favorable also found the assessment to have direct benefits to decision making and intervention strategies. Thus, the authors believed that teachers who became involved with CBA were more likely to participate in developing interventions and evaluating the impact of such strategies. It was stated, then, that because CBA was considered a socially valid tool, its marketability is therefore enhanced which could result in mainstream usage. This could ultimately lead to better development of interventions with the teachers and evaluation of the interventions.

Allinder and Oats (1997) investigated whether students' results were affected by their teachers' acceptability of CBM. In this study, twenty-one teachers each monitored two students in CBM over a four month period. These teachers were administered rating scales which were used to place them into one of two groups: high or low acceptability. Results of this study indicated that teachers who rated CBM as highly acceptable used the program with a greater degree of fidelity. Teachers who showed high levels of acceptance for CBM set more ambitious goals for the results of the CBM from their students than those teachers with lower levels of acceptance. In other words, those teachers who reported "liking" the CBM tools were the same teachers who had higher expectations for the progress of students in their class, when compared to the teachers who "liked" CBM less. In addition, the students of teachers who found CBM to be useful also had greater fidelity and a greater degree of progress by students on CBM throughout the four month period.

It can be surmised from this study by Allinder and Oats (1997) that greater acceptability will likely lead to increased fidelity to the CBM protocol. When there was a high level of fidelity, a student’s progress was more clearly illustrated. It was possible to determine whether a student had positively responded to instructional changes by the teacher. This was beneficial to teachers who wanted to determine which teaching style was most advantageous to increase the skill level of the students in their classes, and create a more positive educational experience for both the teachers and the students.

When a measure is perceived as unacceptable, there are a number of consequences (Eckert et al., 1999). Eckert, Hintze, and Shapiro (1999) noted that within an education setting, unacceptability was often associated with decision-making disagreements and noncompliance with treatment recommendations. These authors also noted that when something such as an assessment measure was not perceived as acceptable, there was less likelihood this measure was used. Therefore, it was believed that the use and effect of assessments measures resulted from the perceived acceptability of the measure (Eckert, Hintze, & Shapiro, 1999).

Assessment Rating Profile

While many different methods may be used to determine the acceptability of a procedure, one such tool that has received widespread use through educational literature is the Assessment Rating Profile (ARP; Chafouleas et al., 2003; Connell, 2006; Eckert et al., 1997, 1999; Eckert, et al.; Shapiro & Eckert, 1992). This tool and its many iterations have been used to gauge perceptions of CBM across many different years (Chafouleas et al., 2003; Connell, 2006; Eckert et al., 1997, 1999; Eckert et al., 1995; Shapiro & Eckert, 1992).

The Assessment Rating Profile was initially developed as a modification and refinement of the 20-item Intervention Rating Profile (IRP; Eckert et al., 1999). The IRP was used to evaluate the acceptability of interventions within the classroom (Witt & Martens, 1983). The adaptation of this rating scale allowed for the communication of acceptability of assessment procedures (Eckert et al., 1999).

In 1995, Eckert, Shapiro, and Lutz used an analog design to determine teacher ratings of CBM when compared to norm-referenced tests. The authors evaluated the acceptability of curriculum-based assessments and norm-referenced tests through the use of the Assessment Rating Profile. General and special education teachers consistently rated curriculum-based assessments as a more acceptable method of assessment. These authors further evaluated the psychometric properties of the ARP, and determined high levels of internal consistency for the rating scale. The Cronbach's coefficient alpha across all 224 participants was .96. This finding suggested a high reliability estimate. In addition, test-retest reliability was obtained using Pearson's product-moment correlation coefficient, and that was found to be .93.

Shapiro and Eckert (1992) administered the ARP to school psychologists to determine the acceptability of curriculum-based assessment. Again, an analog procedure was utilized to compare the ratings of school psychologists towards norm-referenced tests and those towards curriculum-based assessment. The results of this study indicated that while each form of assessment was found to be acceptable by the participating school psychologists, curriculum-based assessment was rated consistently as more significantly acceptable than what is considered "standardized assessment practices" (Shapiro & Eckert, 1992).

Eckert, Hintze, and Shapiro (1997) also used the ARP when assessing school psychologists' perceptions of two types of tools used to evaluate externalizing problem behaviors. School psychologists completed ARP forms after reading a hypothetical case summary. The participating school psychologists rated behavioral assessment measures as more acceptable than traditional assessment measures, such as projective testing.

In an effort to improve the psychometric properties of the ARP, Eckert, Hintze, and Shapiro (1999) revised the original version of the rating scale. The authors noted that previous studies have revealed strong internal consistency of the ARP using Cronbach's coefficient alpha, such as .94 in a study of 500 participants (Shapiro & Eckert, 1994), .96 in a study of 224 participants (Eckert, Shapiro, & Lutz, 1995), and .94 in a study of 339 participants (Eckert, Hintze, & Shapiro, 1997). An exploratory factor analysis was conducted in three separate studies to further examine the scale (Eckert et al., 1997; Eckert et al., 1995; Shapiro & Eckert, 1994). The results in each of these studies yielded similar conclusions. In each of the studies, one primary factor was found within the 18 item scale. The primary factor was labeled "General Assessment Acceptability" and it was believed to uncover the degree to which an assessment was judged as an acceptable method. Two secondary factors were also identified and labeled "Assessment Intrusiveness" and "Assessment Appropriateness" by only one study (Shapiro & Eckert, 1994), while the other two studies only identified "Assessment Intrusiveness" as a secondary factor (Eckert et al., 1997; Eckert et al., 1995). However, it was consistently determined that the ARP was a measure of one strong, unitary characteristic, thus cautioning against the use and interpretation of the secondary factors (Eckert et al., 1999).

Eckert and colleagues (1999) determined that because of the use of a principal components analysis used in the previous investigations, revisions of the scale seemed warranted. They noted that a number of items with poor or bipolar factor loadings were changed to improve the construct validity of the scale. Six items were removed either because of poor wording, irrelevance, or because of an exclusive loading on the secondary factor of “Assessment Intrusiveness.” The Assessment Rating Profile-Revised (ARP-R) was thus developed as a twelve-item, six point Likert scale ranging from Strongly Agree to Strongly Disagree.

An investigation was conducted in order to determine the psychometric qualities of this refined assessment by Eckert and colleagues (1999). The results of this examination using general education teachers (n=201) yielded a strengthened construct validity of the ARP-R. This was determined based on a Cronbach’s coefficient alpha of .99 indicating high reliability estimate. Using Pearson’s product-moment correlation coefficient, test-retest reliability coefficients of .85 (across one-month), .82 (three-month), .84 (six-month), and .82 (twelve-month) were revealed, indicating adequate reliability. In addition, a confirmatory factor analysis specified one underlying latent factor termed “General Assessment Acceptability,” and was found to be measured by each of the individual items on the APR-R.

The Assessment Rating Profile-Revised has also been used in order to determine the acceptability of assessment procedures. Chafouleas, Riley-Tillman, and Eckert (2003) compared the acceptability of three methods for assessing reading: norm-referenced assessment, curriculum-based assessment, and brief experimental analysis. A total of 188 school psychologists were asked to rate one of the three assessment procedures after reading their randomly assigned case description. Results of the ARP-R suggested that the participating school psychologists rated curriculum-based assessment as more acceptable than brief experimental analysis or norm-referenced assessment. The participants indicated that CBA would be more helpful in the development of intervention strategies.

CBM and High-Stakes Testing

One of the distinctive features of CBM is that it is focused on long-term goals rather than short-term objectives (Deno, 2003; Fuchs et al., 1994; Fuchs, 2004; Hintze & Silberglitt, 2005; Hosp et al., 2007; Stecker et al., 2005). Thus, it is possible for a test developer to create an assessment based on what is expected for the students to achieve by the year's end (Deno, 2003; Fuchs et al., 1994; Fuchs, 2004; Hintze & Silberglitt, 2005; Hosp et al., 2007; Stecker et al., 2005). It is then possible to monitor the students' progress because the same objectives are being assessed multiple times across the year. The data compiled through this process allow for curriculum and instruction changes and continuous progress monitoring to determine if the instructional modifications are effective towards meeting those goals (Deno, 2003; Fuchs et al., 1994; Fuchs, 2004; Hintze & Silberglitt, 2005; Hosp et al., 2007; Stecker et al., 2005).

Curriculum-based measurements that monitor a student's progress towards meeting end-of-year goals have a statistical relationship with state-mandated achievement tests (Crawford et al., 2001; Helwig et al., 2002; Hintze & Silberglitt, 2005; Jitendra et al., 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Stage & Jacobsen, 2001). The passage of the No Child Left Behind Act of 2001 (NCLB) mandates that all states must use statewide assessments to demonstrate a district's accountability for student performance (Katsiyannis et al., 2007; NCLB, 2001; Silberglitt et al., 2006). NCLB was intended to spur a national and statewide school reform movement that will challenge the academic standards for all students (Jitendra et al., 2005). This law has been described as “the most expansive role of the federal government in educational matters” (Katsiyannis et al., 2007, p. 160). Schools must be able to demonstrate that the students are making adequate yearly progress (AYP) as measured through a standardized measure of academic achievement, and schools must work towards 100% proficiency of students by the 2013-2014 academic year (Katsiyannis et al., 2007; NCLB, 2001; Silberglitt et al., 2006; Shapiro et al., 2006; Shriberg & Kruger, 2007). In order to determine whether schools are achieving AYP, this law requires that all students are assessed in reading and mathematics in grades three through eight, and from these assessments, the school’s accountability is decided (Katsiyannis et al., 2007; NCLB, 2001; Silberglitt et al., 2006; Shapiro et al., 2006; Shriberg & Kruger, 2007).

Because of the “high-stakes” nature of these tests, school districts have been motivated to find ways of assessing the students’ ability to perform on these statewide achievement tests prior to their actual administration (Jones, 2007; Helwig et al., 2002; Silberglitt & Hintze, 2005). While these tests may not have lasting implications for the test-takers themselves, they can be considered “high-stakes” because of the implications the results have on the educators and schools themselves (Braden, 2007). Thus, the performance of the students on these measures has become a focus for the educators. CBM has been established as a method of monitoring student progress towards reaching long-term goals, which can assist with predicting performance on statewide achievement tests (Deno, 2003; Fuchs et al., 1994; Fuchs, 2004; Hintze & Silberglitt, 2005; Hosp et al., 2007; Stecker et al., 2005). CBM can also provide an advanced notice of students at-risk of failing to meet state standards, and therefore provide teachers with enough time to implement interventions (Helwig et al., 2002). In addition, CBM can provide evidence of adequate yearly progress within a group of students or across a group of students (Helwig et al., 2002). Because of NCLB's emphasis on reading and math, many research studies have investigated the ability of CBM to predict performance on state tests, specifically in these two subject areas (Helwig et al., 2002; Hintze & Silberglitt, 2005; Jiban & Deno, 2007; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005).

A number of studies have determined that "cut scores" may be established to help with prediction of performance on a later measure of achievement (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). A "cut score" would be the score that a student would need to achieve in order for an educator to have a statistical understanding that this student will be able to achieve a passing score on the subsequent high-stakes test, assuming that the student continues to learn at a consistent rate between the time of the CBM and the test (Hintze & Silberglitt, 2005; Silberglitt & Hintze, 2005). The goal of a cut score is to determine whether a student is on track, or progressing at an adequate pace to become proficient in the academic area (Silberglitt & Hintze, 2005).

In order to make predictions regarding a student's likelihood to pass or fail an achievement test, it is important for CBM measures to have diagnostic efficiency statistics (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). There are typically four statistical measures that are calculated to determine the diagnostic accuracy of an instrument, and thus, its ability to predict scores on a separate measure (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006).

The first of these measurements is sensitivity, which refers to the probability that an individual can be positively identified by the diagnostic criteria (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006). This measurement would then indicate the probability of a person failing the achievement test on the basis of the CBM results.

The measurement of specificity refers to the probability that when a diagnostic status is absent, the individual will not be identified by the predictor (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006). This measurement is evident when the probability of those who do pass the achievement test would have been predicted to pass based on the results of the CBM measure is explored.

Positive predictive power is the probability that a student who scored below the cut score on the predictive measure will in fact meet the condition of interest based on the criterion measure (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006). This would mean the probability that a student who was predicted to fail the achievement test on the basis of the CBM cut scores actually failed the achievement test.

Negative predictive power refers to the probability that a student with a score above a predictor cut score does not meet the condition on the criterion measure (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006). This essentially examines the probability that that those who were predicted to pass the achievement test based on a CBM score above the cut score did, in fact, fail the achievement test.

It has been determined through a number of studies using these measures of diagnostic accuracy that CBM can be a strong predictor of performance on various achievement tests in the subjects of reading and math (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Stage & Jacobsen, 2001).

Reading

One of the earliest studies to address the potential use of CBM in predicting performance on an assessment was conducted by Stage and Jacobsen (2001). This study examined the relationship between a measure of oral reading fluency and the Washington Assessment of Student Learning (WASL). In this study, students in the fourth grade were administered oral reading fluency CBMs in September, January, and May, and the slope of these scores was compared to the reading assessment of the WASL using a growth curve analysis. It was determined that the slope was able to predict the May WASL performance, reliably. The authors determined statistically reliable cut scores for each administration of these CBM measures that were able to predict WASL success and failure. The September cut scores were able to determine WASL failure at .41, while they were able to determine WASL success at .90. The authors indicated that with this information, it would be possible to determine which students would benefit from a reading intervention in an effort to alleviate failure on the high-stakes assessment. Thus, it was determined that the measure of oral reading fluency had a great potential for positively affecting performance on a test of reading achievement.

Crawford, Tindal, and Stieber (2001) also investigated how reading CBM can be used to monitor a student’s progress throughout the academic year prior to the use of high stakes standardized assessment. Crawford and colleagues (2001) followed a cohort of second and third grade students over two years. These students were members of blended second- and third-grade classrooms in rural schools districts in Oregon. The majority of the student participants were in general education, that is, 42 of 51, while the remaining nine received special education services. The relationship between oral reading rate (one form of CBM), and scores on statewide achievement tests in reading and math were analyzed. The students were administered three different oral reading probes one day in January. The average of these scores was then compared to the reading and math multiple-choice assessments administered in the state of Oregon.

The results of this study by Crawford and colleagues (2001) revealed a strong relationship between reading rate, as measured through CBM, and performance on both math and reading standardized assessments. The number of words read correctly per minute on the third grade probes was found to have a .60 correlation with the statewide reading assessment and a .46 correlation with the statewide math assessment. The number of words read correctly per minute on the second grade probes was found to have a .66 correlation with the statewide reading assessment and a .53 correlation with the statewide math assessment. Thus, CBM was found to be an important measurement tool that can provide information about students' current and future performance within specific subject areas. The authors concluded that while they do not believe that CBM should replace the standardized assessment procedures, they suggested that CBM could enable a teacher to better utilize the outcomes of such criterion-based assessments on future curriculum.

Another study investigated the predictive validity of CBM procedures with the Michigan Educational Assessment Program's (MEAP) fourth grade reading assessment (McGlinchey & Hixson, 2004). McGlinchey and Hixson (2004) replicated the study by Stage and Jacobsen (2001) using a different state's high-stakes assessment. The participants in this study were administered a one-minute oral reading fluency probe two weeks prior to the administration of the MEAP. It was determined that a moderately strong relationship existed between oral reading rates and performance on the MEAP. The results of this study were able to define cut scores that could be used to accurately predict performance on the MEAP. Based on these results, the authors indicated that these results “link an effective research-based practice to one of the political pressures affecting school districts today” (2004, p. 200). They stated that CBM can be used to assist schools by improving instruction to help students prepare for state assessments.

Further investigation of the predictive validity of reading CBM as it is related to a different state’s achievement test was conducted by Hintze and Silberglitt (2005) in Minnesota. In this study, a comparison was made between different statistical and methodological approaches used to determine cut scores using reading CBM and performance on the Minnesota Comprehensive Assessment (MCA). Students’ performance on R-CBM was obtained through the first, second, and third grades. Students were administered an R-CBM probe during three different intervals across these three years: Fall, Winter, and Spring. These scores were then compared to performance on the MCA at the end of the third grade. The results of this study indicated that R-CBM was strongly associated with performance on the MCA. R-CBM was determined to be accurate and efficient in predicting those students who are likely to pass the reading portion of the MCA. While it was possible to predict performance on the third grade achievement test even from the results of first grade R-CBM, the authors did indicate that there was a stronger correlation between R-CBM measures administered closer to the administration of the MCA.

This study by Hintze and Silberglitt (2005) also determined the best possible statistical means for setting standards and determining cut scores. The authors asserted that each of the three methods of analysis explored (discriminative analysis, logistic regression, and receiver operator characteristic) were able to yield cut scores with adequate levels of diagnostic accuracy and efficiency. The authors indicated that this suggested that R-CBM is a predictor of MCA performance. In addition, the results of this study indicated that these cut scores could be used in school wide screening efforts to compare students to each other and potentially identify those students in need of further remediation within a three-tiered RtI model.

A study by the same authors, Silberglitt and Hintze (2005), added a fourth statistical measure to determine the best possible method for determining cut scores on R-CBM measures in order to predict performance on the MCA. This study also used data from students in first through third grades and compared results from three rounds of reading CBM each year to results on the MCA administered at the end of third grade. However, in this study, students from five different districts were participants. It was determined that R-CBM was a strong tool to predict success on the MCA. The results determined a moderate to high degree of predictive and concurrent validity, and a moderate to high degree of diagnostic accuracy.

The four methods investigated by Silberglitt and Hintze (2005) for determining cut scores were: discriminant analysis, equipercentile method, logistic regression, and receiver operating characteristic (ROC) curve analysis. The results indicated that logistic regression and ROC curve analysis were the strongest methods for determining cut scores. However, the authors indicated that ROC curve analysis provided the most flexibility in establishing levels of diagnostic accuracy, and thus recommended its use within districts.

A later study by Silberglitt, Burns, Madyun, and Lail (2006) analyzed the relationship between two different reading CBM measures and state accountability tests in order to determine if the strength of the relationship changes as a function of grade level. Students from five different districts in the third, fifth, seventh, and eighth grades were administered a measure of reading fluency and/or a Maze reading probe, as well as the corresponding grade level's MCA. Students in grades 7 and 8 were administered both a measure of reading fluency and a Maze assessment, while the students in grades 3 and 5 were only administered a measure of reading fluency.

The results of the study by Silberglitt and colleagues (2006) indicated that R-CBM scores were significantly related to the MCA using Pearson product-moment correlations, with all coefficients meeting or exceeding .50 and significant (p<.001). The results of this study did not find any significant difference between the coefficients associated with the measure of reading fluency and Maze for either the seventh or eighth grade. Neither was more strongly correlated with the state accountability test. It was determined that the relationship between R-CBM and the MCA declined significantly with each advancing grade level. The authors noted that this seemed to be a reasonable finding considering that reading instruction focuses on more complex instructional domains rather than on basic skills as students' progress through their academic careers. It was concluded that R-CBM may not have as great a value in predicting statewide achievement test scores in later grade levels.

Mathematics

While there has been a great emphasis on reading progress in the literature, there has been less research investigating the predicative and concurrent validity of math CBM with state accountability tests in math (Jiban & Deno, 2007; Thurber et al., 2002). There are three distinct constructs of mathematics that has emerged in the education literature: early numeracy, computations, and applications (Hosp et al., 2007; Jiban & Deno, 2007; Thurber et al., 2002). The construct of computations is the completion of math problems when students must know the concepts, strategies, and facts, while the construct of applications is the use of math concepts to solve problems (Hosp et al., 2007; Jiban & Deno, 2007; Thurber et al., 2002). Effective problem solving, then, involves a functional combination of both the computation skills and the knowledge of applications (Thurber et al., 2002).

While reading fluency has been found to be an important component to the development of higher-level reading skills, it is questionable whether math computation fluency is as important to the development of higher-level math skills (Foegen & Deno, 2001). Christ, Scullin, Tolbize, and Jiban (2008) reviewed existing research about computations forms of math curriculum-based measurements in order to summarize relevant findings and discuss implications for practice. These authors noted that math computations CBM has been demonstrated as a reliable and valid set of measurement procedures yet, it should be indicated that these procedures should only be used in specific applications within specific contexts. The authors indicated that CBM focusing on computation skills should only be used to make screening-type decisions because of its limited scope of math achievement. Because of these limitations, the authors concluded that computations assessments alone do have a great deal of predictive utility. Thus, it was noted that additional research may be necessary to improve the procedures and eventually the utility of computations forms of math curriculum-based measurement.

The National Council of Teachers of Mathematics noted in 1989 that math curriculum should focus less on memorization and more on conceptual understanding of mathematics procedures (Jiban & Deno, 2007; McComas, 2000). The emphasis of many studies has been on the use of concepts and applications CBM measures to monitor a student's progress towards meeting a criterion (Helwig et al., 2002; Jitendra et al., 2005; Shapiro et al., 2006). Overall, math concepts and application CBMs have been shown to be predictors of achievement on state accountability tests in math (Helwig et al., 2002; Jitendra et al., 2005; Shapiro et al., 2006).

Helwig, Anderson, and Tindal (2002) used a measure of math that focused on concepts, as opposed to fluency, in order to determine whether such a brief measure would correlate with scores on a computer adaptive test (CAT) that was developed with a western state to approximate the official statewide math assessment. The authors' choice to utilize a CBM more grounded in math concepts, as opposed to math computations was as a result of a call by the National Council of Teachers of Mathematics (NCTM) to emphasize conceptual understanding of mathematical procedures and algorithms, instead of memorization (McComas, 2000). The authors suggested that CBM tasks with more emphasis on conceptual understanding may be a better predictor of overall math achievement because knowledge of basic concepts is necessary to successfully apply this knowledge to a higher-level question, and because of the overlap that exists between various domains within mathematics. The scores obtained from general education students were compared to scores obtained from students with learning disabilities in order to determine whether there would be any difference in the way CBM should be used with the different student populations.

Helwig and colleagues (2002) concluded that the CBM tasks used were effective in predicting scores on a statewide mathematics assessment for the whole group (.60) and individual general education students (.53). However, the tasks were only moderately effective in predicting scores for students with identified learning disabilities (.38). In addition, using a discriminant function analysis (DFA), the authors were able to predict which students would meet the state mathematics standards with 87% accuracy. They also noted that students with learning disabilities scored lower on both the CBM task and the CAT. The authors indicated that the CBM instrument may not have been sensitive enough to measure the progress of these students effectively enough to monitor their typical growth within the course of the semester.

A later study by Jitendra, Sczesniak, and Deatline-Buchman (2005) investigated the validity of word problem-solving tasks as indicators of mathematics proficiency as defined by the Mathematics Computation and Concepts and Applications subtests of the TerraNova. The participants included 77 third grade students who were administered a number of assessments in order to understand the math ability of the students. An abbreviated battery of the Stanford 9 was administered, which included a test of Problem Solving and a test of Procedures. This test was administered in December of the third grade year. The TerraNova achievement test was administered to the students in April of their third grade year. The students also completed Word Problem-Solving Fluency (WPS-Fluency) probes every other week for a 16-week period of the study. They were given 10 minutes to complete eight problems on this measure. Finally, students were administered one computation fluency probe each week during the four weeks of the winter session and four weeks of the spring session. The students were provided with three minutes to complete 25 problems.

The results of this study by Jitendra and colleagues (2005) indicated that the word problem-solving CBM probes in both the winter and the spring created moderate correlations with most of the other measures used in the study. There were moderate to high correlations with the Stanford 9 Problem Solving subtest (winter = .71, spring = .54), the Concepts and Applications subtest of the TerraNova (winter = .69, spring = .58), and the Computation subtest of the TerraNova (winter = .62, spring = .48). There were moderate to low correlations with the Procedures subtest of the Stanford 9 (winter = .58, spring = .38). It was determined that the CBM problem-solving measure was a moderate to strong predictor of mathematics competence, but it was stronger when combined with the computation CBM measure. The authors noted that when time constraints prohibit the administration of both types of probes, the problem-solving CBM may provide adequate information about math competence because "solving word problems requires integrating and applying both concepts and procedures" (p. 368).

In order to further explore CBM as it relates to the Pennsylvania statewide achievement test, Shapiro, Keller, Lutz, Santoro, and Hintze (2006) assessed general education students in two school districts in eastern Pennsylvania. CBM measures in reading, math computation, and math concepts and applications were administered to students in the third, fourth, and fifth grades at three different intervals during the academic year, and the results were then correlated with various standardized assessments. The students in the third and fifth grades in both districts were administered the Pennsylvania statewide achievement test, the Pennsylvania System of School Assessment (PSSA) in either March or April. Fourth grade students in the first district were administered the MAT-8 while fifth grade students in this district were administered the Stanford Diagnostic Reading Test (SDRT). Fourth grade students in the second district were administered the SAT-9.

Results of the study by Shapiro and colleagues (2006) indicated that there was a moderate to strong correlation between the MAT-8 and the CBM reading scores, ranging from .519 to

.724. There were somewhat lower correlations found between the CBM measures and the SDRT: .524 in the fall, .518 in the winter, and .551 in the spring. There were consistent moderate to strong relationships obtained between CBM and all of the subtests of the SAT-9. The PSSA scores were correlated with the results of the CBM measures for the fifth grade students only, because at the time of the study, performance categories had only been established for the fifth grade. In addition, because the results of the winter CBM measures correlated most strongly with the PSSA scores, only these scores were analyzed. Results indicated that the CBM reading measures were moderately to strongly correlated with the PSSA (.70), and a moderate correlation was found between math computation and PSSA scores, ranging from .50 to .53. Moderate correlations were also found between the math concepts and applications CBM compared to the PSSA scores, ranging from .61 to .64. The authors concluded that CBM assessments have strong predictive power that may help identify students who are likely to meet or exceed the proficient criterion on the PSSA. The results of this correlation could lead to designing short-term remediation programs that focus on the skills needed to be successful on statewide assessments.

Another study completed by Jiban and Deno (2007) assessed the technical adequacy of three different kinds of CBM used with third and fifth grade students as they related to the Minnesota Comprehensive Assessment (MCA) math test. The measures used with the 38 third grade and 55 fifth grade participants in this study were two different math CBM measures and one reading CBM which were administered within two weeks of the administration of the MCA. There was first a measure of basic math skills organized in a horizontal, as opposed to vertical, format. Students were provided with one minute to complete as many problems as possible. A cloze measure of math facts was also used. That is, math problems were again fashioned in a horizontal fashion, but students were requested to fill in the missing number in any of the three positions in the problem. The students were again given one minute to complete as many problems as possible. A maze reading measure was also used where the students were asked to read silently and circle one of three choices each time they came to a missing word in the sentence. The students were again given one minute to complete as many sentences as possible. These results were then correlated with the criterion measure, the MCA math assessment.

The results of this study by Jiban and Deno (2007) revealed moderate correlations between the MCA and all three CBM measures used in both grade levels. Two different scoring methods were used in each of the math CBM tools in order to determine whether one method would be more effective. The two methods were the number of problems correct and the number of problems correct minus the number of problems incorrect. The maze reading measure was only scored as correct answers minus incorrect answers. Because of the small sample size of third grade students, there were problems with reliability at this level, thus only limited conclusions could be made about this sample. The basic math facts CBM at this level did not reveal significant results; however, the other two measures used with the third grade students did reveal moderate correlations with the MCA. The cloze math CBM revealed correlations of .38 and .44 in the two scoring methods respectively, and the maze reading correlation with the math assessment of the MCA was .44. The fifth grade basic math facts correlated with the MCA were .55 and .57, while the cloze math facts measure results were both .59. The maze reading measure correlated with the math MCA assessment at .63. The authors indicated that the cloze math CBM was better able to predict scores on the MCA than the traditional math facts for the fifth grade students. Because of these results, the authors indicated that the cloze math procedure should be more widely considered not only because of its predictive validity but also its simplicity and brevity of administration and creation. In addition, because of the strong correlations found in both reading and math and the importance of each skill in meeting the criterion on state standardized tests, the authors suggested using each of these measures in tandem to obtain the most information about the students. The authors noted that together, the cloze math facts and maze reading measures explained 52% of the variance for the fifth grade students.

One of the most influential studies that monitored student progress using a CBM focusing on concepts and applications was completed by Fuchs and colleagues (1994). The authors were concerned that only limited portions of math achievement were being addressed by the CBM measures used at the time. Thus, the authors determined technical features of a math CBM that would address number concepts, counting, applied computation, geometry, measurement, charts and graphs, money, and problem solving.

In order to answer this question, 140 students in second, third, and fourth grades were administered weekly measurements throughout the school year. These computer-assisted CBM probes addressed the domains of concepts/applications and computation. The students were given limited time on each set of probes. For the concepts and applications probes, their time depended on grade level: second graders were given eight minutes while the second and third graders were given six minutes. On the computations probes, the second graders were provided with 1.5 minutes, the third graders were provided with two minutes, and the fourth graders were provided with three minutes. With the assistance of a scoring program, the results provided information about mastery of specific skills on the concepts and applications probes, in addition to the overall score of total number of points correct. The scores were then graphed so that student progress could be monitored. The computation probes were scored as total number of correct digits. These scores were then compared to Normal Curve Equivalent (NCE) obtained by the students on the Comprehensive Test of Basic Skills (CTBS) which was administered to the students by the school district in the spring of the academic year.

In order to determine the effectiveness of the concepts and applications probes, a number of analyses were completed. By completing a least-squared regression between calendar days and CBM scores, the daily rate of increase in performance was completed for both computations and concepts and applications. The weekly slope obtained for computations scores ranged from .25 to .70, and the weekly slope obtained for concepts and applications ranged from .40 to .69. When comparing the slope results of the older grades, where the impact of reading the concepts and applications probes had less of an effect, the progress of students in both computations and concepts and applications appeared to be parallel. This suggested to the authors that instruction in one area of mathematics (i.e., computations) may encourage improvement in the other area of mathematics (i.e., concepts and applications). The results of this analysis also revealed high levels of internal consistency for the concepts and applications probes, ranging from .94 to .98, as measured through a correlation comparing scores obtained on varying days throughout the year. These results indicated that these CBM can provide reliable results for teacher use when monitoring student progress. When assessing concurrent validity of the concepts and applications CBM scores to the results of the CTBS, the correlations ranged from .63 to .81. Thus, there appeared to be a strong relationship between student performance on the CBM probes and the CTBS, which was described as “a well-accepted and commonly used measure of mathematics proficiency” (p. 37).

This study, overall, provided an important contribution to the research about curriculum-based measurement in mathematics. This was the first documentation that student growth could be assessed through a CBM consisting of more comprehensive mathematics curriculum. In addition, the technical adequacy of a CBM focusing on math concepts and applications was supported. The authors recommended that further research be conducted to explore the feasibility of these assessment tools and their impact on overall student progress.

To further the research of Fuchs and colleagues (1994), Connell (2006) developed math CBM concepts and applications probes © that replicated the earlier research and extended the validity of these types of probes to different criterion assessments. In Connell's study, 173 students in grades one through five were administered a single skill computations CBM, three multiple skill computations CBM, a concepts and applications CBM, a maze reading CBM, the Iowa Test of Basic Skills (ITBS), the Louisiana Education Assessment Program for the 21st Century (LEAP), and the Woodcock-Johnson III (WJIII) math subtests. The goal of this study was to determine which type of math CBM would have the strongest correlation with these commonly used criterion referenced assessments.

The concepts and applications CBM probes © were developed for use in this study by Connell (2006). In creating these probes, the two main constructs of math, that is, computations and applications, were included (Thurber et al., 2002). The math curriculum from seven randomly chosen states was compared to the National Assessment of Education Progress as well as the "Nation's Report Card" to determine the six common content standards that would be addressed in the probes. The six standards that were included in these probes were a) number sense, b) data analysis, c) patterns, d) algebra, e) measurement, and f) geometry. Six problems from each of the six standards, with 36 problems total, were included in the probes at each grade level.

The concepts and applications probes, along with the other unique measures, were administered to the students and scored by psychology graduate students who had been trained in the administration of the standardized assessments as well as CBM procedures. The typical classroom assessments, such as the ITBS and the LEAP, were administered according to school protocol by the classroom teacher and other school personnel. Following the administration of these measures, the classroom teachers were asked to provide their perspectives about the CBM tools through the use of the ARP-15.

This study replicated the results of previous studies investigating computations CBM when compared to criterion referenced assessments. That is, moderate to strong correlations were found between the multiple skill and single skill computations probes when compared to the computations fluency assessments on the ITBS and the WJIII. However, when comparing the overall math scores obtained on the various criterion assessments, the concepts and applications probes were found to correlate better than either of the computation CBM probes. There was less variation in scores in the results obtained by the students on the concepts and applications probes than the computations probes. In addition, the classroom teachers reported higher ratings of the concepts and applications probes than the computations probes.

The results of the study by Connell (2006) suggested that the traditionally used progress monitoring computation CBM tools may not be as effective at when comparing performance or as preferred by teachers as a measure of concepts and applications. However, limitations to the study by Connell (2006) left some questions unanswered about the effectiveness of these probes to student progress over time as well as the correlation of the probes to other criterion assessments used in other schools.

Overall, CBM has been found to be valid and effective tool for monitoring student progress and predicting performance in various content areas (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005). Some forms of CBM correlated higher with commonly used criterion assessments that are often used to make school wide, high-stakes, decisions (Connell, 2006; Foegen & Deno, 2001; Foegen, Jiban, & Deno, 2007; Francis et al., 2005; Fuchs & Fuchs, 2002; Fuchs & Fuchs, 2007; Helwig et al., 2002; Hintze & Silberglitt, 2002; Hosp & Hosp, 2003; Jiban & Deno, 2007; Jitendra et al., 2005; McCardle et al., 2008; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005; Stecker et al., 2005). Thus, it is important to continue to research these processes in order to determine the most effective methods of prediction to influence instruction and ultimately impact performance on high-stakes testing.

The purpose of the current study was to replicate the study completed by Connell (2006) which produced significant results comparing concepts and applications probes to various criterion assessments. However, this study extended the findings of Connell (2006) by using the validated concepts and applications probes as progress monitoring tools as well as possible predictors of performance and comparing these results to scores obtained on commonly used multiple skill computations probes. In addition, the concepts and applications probes were compared to a different criterion assessment to extend the understanding of the effectiveness of these probes to predict performance. This study also investigated the beliefs of teachers about the various CBM probes before and after use when the teachers themselves administered and scored the probes. It was the overall goal of the study to find an effective, yet highly regarded assessment that could positively influence math instruction by predicting performance on high-stakes, end of year assessments in a variety of locations.

CHAPTER 3

METHODOLOGY

Participants

Experimenter

The experimenter worked for a private nonprofit educational organization that provided a full range of services to regular education and special education kindergarten through twelfth grade students in parochial and private school systems in a mid-Atlantic state. The services provided included counseling, crisis intervention, and psychological-educational evaluations; educational services, such as corrective reading and math instruction, speech and language therapy, and enrichment programs; and special services, such as homebound instruction and professional development services.

Teachers

Fifteen corrective education teachers also working for the non-profit organization were asked by the experimenter to participate. They were encouraged to request that their respective schools use the math progress monitoring program within the second and third grades of the schools in which they worked. The corrective education teachers worked at multiple nonpublic schools in a large urban school district. They provided regular remediation services in math and/or reading to forty-four private schools. All corrective education teachers had a Master's Degree in Education or a related field, or at least the credit equivalent to the degree. One of the participants was a Reading Specialist. The corrective education teachers ranged in experience from thirty to forty years. Only eleven of these fifteen were able to encourage a school's participation in the program. In addition, two school psychologists encouraged participation in schools where they worked. Only nine corrective education teachers, who were present during the initial meeting describing the purpose of the study, were asked to complete the Assessment Rating Profile, and fully participate in this study.

Students

A total of 44 schools were asked to participate in the study. Only 21 schools chose to participate. Sixteen of these schools were parochial schools with tuition beginning at $2,000 for the 2007-2008 academic year. The remaining schools were private schools affiliated with a religious organization. The tuition of these schools ranged from $2,000 to $9,000 for each student during the 2007-2008 academic year.

Within these 21 schools, 19 second grade classrooms with a total of 453 students participated in the study as well as 16 third grade classrooms with a total of 371 students. Not every classroom teacher chose to utilize the applications probes. Therefore, of the 19 second grade classrooms that participated, only 13 of those classrooms administered the applications probes. Of the 453 second grade students that completed computations probes, only 252 of these students were also administered the applications probes. In addition, of the 16 third grade classrooms that participated, only 12 of those classrooms administered the applications probes. Therefore, of the 371 third grade students that completed the computations probes, only 215 of these students were administered the applications probes. The class sizes of these participating classrooms ranged from 5 – 35 students per classroom. The classrooms were diverse in both ethnicity and socio-economic status.

Twenty-five of the 26 participating parochial school classrooms provided TerraNova scores to compare to the CBM results. There was one participating classroom that had chosen not to take the TerraNova assessment during that academic year. Thus, TerraNova scores were obtained for 394 second grade students (all of whom completed computations probes while only 193 completed applications probes) and 286 third grade students (all of whom completed computations probes while only 130 completed applications probes).

Measures

Four primary measures were used in this study: two forms of math curriculum based measurements (CBM), a social validity measure, and a state administered assessment.

Math Curriculum Based Measurements

Two separate types of math CBM probes were used in this study. The assessments (called probes) were based on the state's math curriculum as well as the math curriculum for second and third grade developed by the local parochial school education board. Thus, these probes addressed the goals established by these curriculums for the end of the year in the second and third grade.

Computation Fluency Probes

Multiple skill computation probes were administered to the second and third grade students. The computation measures were fluency-based, meaning that students had a limited amount of time to complete as many problems as they could. Second grade assessment time was two minutes and the third grade was three minutes. The probes were developed two different ways. The second grade probes were created using an assessment generator application from interventioncentral.org. This website provides an application specifically designed to build math computation assessments in accordance with the procedures described by Fuchs (1994). These probes can be found in Appendices D through F.

In developing the third grade probes, the Monitoring Basic Skills Progress (MBSP) program published by Pro-Ed, Inc.© was used This series has a great deal of empirical data supporting its reliability, criterion validity, and sensitivity in detecting growth in students at the elementary level (Fuchs, Hamlett, & Fuchs, 1998; Foegen et al., 2007). From this series, six math worksheets were selected as clearly representing the end-of-year curriculum goals. These six worksheets were then divided into a total of three probes. These probes are not included in the Appendices to respect the author's copyright.

Three alternate-form probes were generated for each grade (second and third) to be administered at three different points in the academic year.

Concepts and Application Probes

Second and third grade applications probes were developed by Connell (2006) ©. The applications probes were developed according to the guidelines established by the National Assessment of Education Progress (NAEP) and National Council of Teachers of Mathematics (NCTM). They contained six main content areas: number sense, patterns, data analysis, geometry, measurement and algebra. Each grade-level probe contained six problems from each content area, thus 36 problems total. The probes were developed from a commercially used math curriculum (Houghton-Mifflin; *Math Central*©) using the following procedures. First, a table of specifications was developed for each of the six content areas. Then 10 items from each content area were selected from the Math Central unit tests that matched the table of specifications. Next, items where placed under a content area in the table. Then, six of the items were selected using a quasi-random selection process (items selected grew increasingly more difficult as the unit tests moved from the introductory level to the advanced level). The items were modified to protect the publisher's copyright and reviewed by an expert in curriculum and instruction. These probes, which have been copyrighted, are included in Appendices G and H.

Assessment Rating Profile – 15

The Assessment Rating Profile – 15 required teachers to respond to 15 items on a six-point Likert scale where *one* indicated that the respondent "strongly disagree[d]" with the statement, and *six* indicated that the respondent "strongly agree[d]" with the statement. Only one underlying latent factor termed "General Assessment Acceptability" is measured by each of the individual items on the APR-15. The survey is included in Appendix I and J.

TerraNova, Second Edition

TerraNova, Second Edition, (TN2) was a national, norm-referenced achievement test published by CTB/McGraw-Hill. This specific version of the test was part of a series that has been revised and expanded since the first version, published in 1997. It measured concepts, processes, and skills that were taught in the subject areas of reading and language arts, mathematics, science, and social studies (Cizek, 2005; Stevens & Zvoch, 2007). The TerraNova series was used by the state to compare students' scores with other students across the country.

The TerraNova used a multiple-choice format that assessed students in the aforementioned five areas. This test provided results in traditional forms of norm-referenced scores including national and/or local percentile ranks, grade equivalent scores, and normal curve equivalent, as well as more unique norm-referenced scores including a developmental scale score which can be used to track student progress across grade levels (Cizek, 2005). The national standardization of this test, or the norming, occurred during the fall of 1999 and the spring of 2000 (Johnson & Mazzie, 2005). The TN2 also provided information about student performance by categorizing according to one of five labels: Step One, Progressing, Nearing Proficiency, Proficient, and Advanced (Cizek, 2005). These labels were similar to those outlined by the National Assessment of Educational Progress: Basic, Proficient, and Advanced (Cizek, 2005).

This test was found to have high levels of reliability, according to interrater agreement indices and weighted kappa coefficients (Cizek, 2005). Internal consistency was reported to range from the high .80s to the high .90s (Johnson & Mazzie, 2005). The TerraNova reported that it has strong content validity. They noted that efforts were made to coordinate the graphics and the content of the test with student and teacher questionnaires indicating the content of curriculum used in the classroom (Johnson & Mazzie, 2005). Thus, the correspondence between the test content and instructional content was increased (Johnson & Mazzie, 2005). Very little evidence was provided indicating construct validity (Cizek, 2005), but it is indicated that there is evidence of convergent and discriminant validity (Johnson & Mazzie, 2005). That is, the test developers did indicate that tests designed to measure similar skills, such as Reading and Language Arts, correlated higher than tests designed to measure different skills, such as Reading and Mathematics (Johnson & Mazzie, 2005).

Overall, reviewers of the TN2 indicated that it is a well-developed instrument with evidence of reliability and validity as a measure of student achievement in skills that are important and relevant in the grades and subjects assessed (Cizek, 2005; Johnson & Mazzie, 2005).

Procedure

In order to investigate the reliability, validity, and acceptability of these CBM computation and applications probes, the following procedures were utilized. At the beginning of the 2007-2008 academic year, the CBM probes were presented to the staff of the private nonprofit educational organization. The brief presentation consisted of an overview of the program and a summary of the results that can be yielded from the use of the program within the classroom. Approximately one week following this initial presentation, a training session was held for the corrective education teachers. This hour long session was conducted by the group of five teachers and school psychologists who had worked on developing the math progress monitoring program. Within this training, many topics were discussed, including the research behind progress monitoring, the need for the program, and how to successfully implement the program within the schools. The primary presenter was the investigator, a doctoral student in school psychology.

All of the necessary materials were made available to each teacher through the Information Technology Consultant at the private nonprofit educational organization. Thus, the information could be obtained by each corrective education teacher and placed on his or her flash drive making it accessible from any location at any time. Included on this flash drive were a brief manual outlining the basic concepts of curriculum based measurement and progress monitoring, administration instructions, Microsoft Excel graphing templates, multiple skill computation probes, and single skill probes.

Each corrective education teacher at the private nonprofit educational organization was encouraged to find at least one classroom teacher willing to participate in the program. They were provided with a brief flyer encouraging the school principal to consider allowing this math progress monitoring program into the school.

After the corrective education teachers reported back to their supervisor relating how many classroom teachers were interested in participating, probes were copied for the teachers use. The correct number of applications and computation probes were provided for the corrective education teachers to deliver to the school for each round of the assessment. In addition, copies of the administration scripts and answer sheets were provided to the teachers for their use for each round of the assessment.

Prior to the administration and scoring of the initial probes in October of 2007, a brief meeting was held for the corrective education teachers of the private nonprofit educational organization. At this meeting, the consent form and Assessment Rating Profile-15 were distributed to the nine corrective education teachers. A brief overview of the research project was provided, and the individual corrective education teachers were asked to sign and return the consent form and complete the rating scale. Both of these materials were to be returned to the investigator as soon as possible. The corrective education teachers that were not present at this meeting were not provided with these materials and thus were not asked to participate. All of the rating scales were returned to the investigator by the beginning of November.

The corrective education teachers worked in concert with the classroom teachers to determine an adequate time to conduct the October assessment. After the assessment date was determined, the materials were provided to the classroom teacher. The corrective education teachers were encouraged by their supervisors at the private nonprofit educational organization to be present during the administration of the assessment. However, if scheduling conflicts arose, the classroom teacher was provided with a script that was to be followed exactly to ensure fidelity in the administration process. The administration scripts used by both the classroom teachers and the corrective education teachers are included in Appendices A through C.

After the probes were administered to each student in the classroom who had returned a permission slip, according to the policy and procedures of the private nonprofit educational organization, the probes were returned to the corrective education teachers. The corrective education teachers scored the probes in two different ways. The computations probes were scored as digits correct per minute (DCPM) while the applications probes were scored as percentage of total problems correct. If the corrective education teachers needed additional assistance throughout the scoring process, they were encouraged to contact the investigator either on the phone or via email to set up a meeting.

Following the scoring of the probes, the corrective education teachers entered the scores from the computation probes in an accompanying Microsoft Excel database. Training on the use of this database was provided during the initial training session held in September. However, if the corrective education teachers continued to have questions, they were again encouraged to meet with the investigator to obtain assistance.

The database used for the computations probes provided two charts for each classroom and three graphs for second grade classrooms and five graphs for third grade classrooms. These visual demonstrations of the results were provided to the classroom teachers to illustrate the progress of the students in these basic skills throughout the three administrations of the probes.

The applications probes were given to the investigator who entered the scores in a separate database which also provided a visual demonstration of the results. Two charts and seven graphs were obtained through the applications database and were provided to the remediation teacher who then passed the results to the classroom teacher. The graphs and charts that were provided to the classroom teacher were meant to incite a consultative relationship between the classroom teacher and remediation teacher in order to benefit the math curriculum within the classroom and further identify students in need of further remediation in basic math skills.

The process of probe administration and analysis of scores was completed a total of three times throughout the academic year. While alternate forms were developed for each administration of the computations probes, the same version of the applications probes were used during each administration. After the second and third administrations, the corrective education teachers and two school psychologists were provided with the option of having the Information Technology Consultant enter the scores for each of their students into the Microsoft Excel database. Most of the corrective education teachers chose to use this option, and thus provided the scored computations probes to this person, who then entered the scores into the database. The investigator continued to enter all of the applications probes for those schools participating with the applications portion of the program.

The probes were administered to the second and third grade students in October, January, and May as a universal screening tool. After each administration, the same number of charts and graphs were provided to the classroom teachers for feedback. However, with each administration, an additional bar was added to the graphs. Thus, it was possible for the teachers to note how much a student progressed with each administration throughout the year.

In May, prior to administering the final sets of probes, the teachers were asked to again complete the ARP-15. The same rating scale was used for this second administration as was completed at the beginning of the academic year. However, for this administration, the teachers were asked to complete the rating scale with a specific set of probes in mind. Thus, each teacher who utilized the computations probes were asked to complete an ARP-15 considering only the computations probes. Each teacher who used the applications probes was asked to complete an ARP-15 considering only the applications probes. Examples of these probes were attached to the rating scale when they were distributed so that the teachers might be able to refer back to the probes that were used when relaying their opinions. These rating scales were distributed via the participants' office mailboxes in the middle of May. Teachers were asked to complete the rating scale immediately and turn in the papers to the investigator by the end of the day of their reception.

After the results of the final set of probes was provided to the classroom teachers, the corrective education teachers and school psychologists obtained the results of the most recent administration of the TerraNova assessment either through the classroom teacher or the school's administration. These results were then paired with the results of the probes, and the students' names were removed from the results.

Data Analysis

In the current study, it was hypothesized that the different CBM measures could be used to predict student performance on the appropriate grade level TerraNova math assessment. In order to determine the validity of this hypothesis, the following analyses were conducted on the data.

Initially, the technical adequacy of the CBM scores was investigated through the use of a Pearson correlation in order to determine the relationship of the scores obtained to each other, and thus reliability of the CBM across the three separate administrations. Each of the scores obtained from the two different CBM measures were compared to the scores obtained by the same students on the appropriate grade level TerraNova through the use of a Pearson correlation and linear regression. In addition, a repeated measures ANOVA was used to investigate the relationship between the various CBM scores obtained through the two different types of probes.

Additionally, it was hypothesized that the results of the Assessment Rating Profile-15 would reveal significantly more favorable beliefs and attitudes towards the use of CBM in the second administration of the rating scale than the first. It was assumed that the corrective education teachers would report more favorable attitudes towards the applications probes than the computation probes because of a general emphasis on conceptual learning by teachers within the classroom.

In order to investigate these hypotheses, the means obtained through the pre- and post-rating scales will be compared using a nonparametric correlated t-test. These means were then investigated using a Wilcoxon t-test in order to better understand the relationship between the various means obtained through the use of the ARP-15.

CHAPTER 4

RESULTS

In the current study, the following hypotheses were investigated. First, it was hypothesized that the math computation and applications measures could be used to predict student performance on grade level TerraNova math subtests. Second, it was hypothesized that the teachers would prefer the math applications assessment over the math computation assessment as reported by Connell (2006).

The mean scores were obtained for the 133 second grade students who completed all of the computations probes, concepts and applications probes, and the TerraNova math subtests. The same data was obtained for the 108 third grade students with the same scores available. Tables 1 and 2 present the mean scores on both the computations and concepts and applications probes obtained by the second and third grade students, respectively.

Table 1. Mean Scores Obtained by the Second Grade Students

	October	January	May
Computations (DCPM[1])	5.46	10.30	9.33
Applications (% correct)	46.49%	58.63%	65.77%

[1] Digits Correct Per Minute

Table 2. Mean Scores Obtained by the Third Grade Students

	October	January	May
Computations (DCPM)	4.07	10.61	9.88
Applications (% correct)	51.85%	56.35%	64.74%

A repeated measures Analysis of Variance, with sphericity assumed, was conducted in order to determine the students' overall progress across time. Table 3 presents the results of the repeated measures ANOVA. Based on these results, it was determined that the students significantly improved their scores across times in both second and third grade when assessed with both computations and concepts and applications probes.

Table 3. Repeated Measures ANOVA of CBM Scores across Time

CBM Probes	SS	Df	Mean Square	F	Significance	Partial Eta Squared
Computation						
Second Grade	1742.001	2	871.001	35.460	.000	.212
Third Grade	2775.366	2	1387.683	45.714	.000	.299
Application						
Second Grade	25278.119	2	12639.060	119.052	.000	.474
Third Grade	9237.873	2	4618.937	39.799	.000	.271

Paired t-tests were conducted in order to further explore the significance of the student's scores from one administration to the next. The results of the paired t-tests for the results of the computations probes and the concepts and applications probes are presented in Tables 4 and 5, respectively. Based on these results, it was determined that the students significantly improved in most intervals. However, the students in both second and third grade did not demonstrate a significant improvement in scores obtained on the computations probes when comparing January results to May results. Both the second and third grade students scored the highest number of digits correct per minute during the January administration (10.30, 10.61). Both the second and third grade students showed a decline in the number of digits correct per minute from the January administration to the May administration. Thus, the paired t-tests comparing the means of the January and May computations administrations did not yield significant results, as noted in Table 4.

Table 4. Paired t-Test Comparing Means of Computations CBM Administrations

	Mean	Std. Deviation	Std. Error Mean	t	Sig. (2-tailed)
Second Grade					
October – January	-4.83	7.57	.65	-7.36	.000
October – May	-3.86	6.08	.52	-7.32	.000
January – May	.97	7.27	.63	1.54	.125
Third Grade					
October – January	-6.54	8.25	.79	-8.23	.000
October – May	-5.80	5.27	.50	-11.44	.000
January – May	.73	9.28	.89	.82	.413

There was an observable upward trend noted in the means obtained for the concepts and applications probes across administrations, as demonstrated by the results of the paired t-tests presented in Table 5. Both the second and third grade students answered correctly a significantly higher percentage of problems across each administration of the probes. The highest percentage of problems correct for both the second and third grade students was obtained during the May administration (65.77%, 64.74%).

Table 5. Paired t-Test Comparing Means of Concepts and Applications CBM Administrations

	Mean	Std. Deviation	Std. Error Mean	t	Sig. (2-tailed)
Second Grade					
October – January	-12.13%	15.58%	1.35%	-8.98	.000
October – May	-19.28%	16.47%	1.42%	-13.49	.000
January - May	-7.14%	11.07%	.96%	-7.43	.000
Third Grade					
October – January	-4.50%	17.75%	1.71%	-2.63	.010
October – May	-12.88%	14.60%	1.40%	-9.17	.000
January - May	-8.38%	12.92%	1.24%	-6.74	.000

A Pearson correlation was completed in order to determine the relationship of the CBM scores to each other. This was used to examine the reliability of the CBM measures across three separate administrations. The concepts and applications probes were found to have more reliability across administrations than the computations probes in both the second and third grade levels, ranging from .588 to .832. Tables 6 and 7 present the correlations between the scores obtained at each administration of a CBM probe for the second and third grade students, respectively. The strongest correlation was found in the scores obtained from the second grade students when correlating the results of the January concepts and applications probes and the results of the May concepts and applications probes (.832), while the weakest correlation was found in the scores obtained by the third grade students when correlating the results of the January computations probes and the October concepts and applications probes (-.008).

Table 6. Correlations between Various CBM Measures for Second Grade Students

	Computation			Application		
CBM Probes	October	January	May	October	January	May
Computation						
October	--	.486**	.244**	.490**	.396**	.411**
January		--	.553**	.284**	.177*	.244**
May			--	.135	.137	.104
Application						
October				--	.671**	.588**
January					--	.832**
May						--

*p < .05. **p < .01.

Table 7. Correlations between Various CBM Measures for Third Grade Students

	Computation			Application		
CBM Probes	October	January	May	October	January	May
Computation						
October	--	.347**	.471**	.363**	.466**	.417**
January		--	.252**	-.008	.036	-.006
May			--	.364**	.387**	.345**
Application						
October				--	.634**	.724**
January					--	.809**
May						--

*p < .05. **p < .01.

The results obtained through the use of these probes were then compared to the Total Math National Percentile (NP) scores obtained by the students on the TerraNova assessment. A Pearson correlation was completed in order to determine the relationship between the results of the probes on each administration and the NP results of the TerraNova assessment. Table 8 presents the results of the Pearson correlation. A stronger relationship was found between the scores on the concepts and applications probes and the TerraNova NP scores than the scores on the computations probes and the TerraNova NP scores. Correlations with the results of the computations probes ranged from .181 to .456, while correlations with the results of the concepts and applications probes ranged from .609 to .805 (p<.01).

Table 8. Correlation of CBM Probes to TerraNova Total Math National Percentile

	Computation			Application		
Grade	October	January	May	October	January	May
Second	.439**	.284**	.225**	.609**	.723**	.751**
Third	.456**	.181	.351**	.730**	.721**	.805**

*p < .05. **p < .01.

The results obtained through the use of these probes were then compared to the Total Math Normal Curve Equivalent (NCE) scores obtained by the students on the TerraNova assessment. The NCE scores were used in this analysis because they are a standard score and can be more readily compared to the scores obtained on the various CBM measures (Hogan, 2003). A Pearson correlation was completed in order to determine the relationship between the results of the probes on each administration and the NCE results of the TerraNova assessment. Table 9 presents the results of the Pearson correlations. A stronger relationship was again found between the results of the concepts and applications probes and the TerraNova NCE results than the results of the computations probes and the TerraNova NCE results. Correlations with the results of the computations probes ranged from .201 to .510, while correlations with the results of the concepts and applications probes ranged from .613 to .803 ($p<.01$).

Table 9. Correlation of CBM Probes to TerraNova Total Math Normal Curve Equivalent (NCE)

	Computation			Application		
Grade	October	January	May	October	January	May
Second NCE	.510**	.321**	.230**	.613**	.716**	.749**
Third NCE	.496**	.201*	.373**	.723**	.730**	.803**

*p < .05. **p < .01.

A linear regression analysis was completed comparing the results of both types of probes with the Total Math Normal Curve Equivalent from the TerraNova subtests in order to obtain more specific information about the predictive validity of the CBM to the TerraNova subtests. The Normal Curve Equivalent was used to represent the scores from the TerraNova subtests because this measure is considered to be a standard score (Hogan, 2003) and is thus more comparable the scores obtained from the curriculum-based measures. A stronger relationship was again found with the concepts and applications probes than the computations probes, as noted in Table 10.

Table 10. Linear Regression Statistics for CBM Probes and TerraNova Total Math Normal Curve Equivalent (NCE)

	Computation		Application	
Grade	R	r^2	R	R^2
Second NCE	.523	.273	.784	.614
Third NCE	.521	.271	.836	.700

Further analysis was completed in order to understand which assessments, at which times were most predictive of a student's Normal Curve Equivalent score on the TN2. While analyzing and interpreting the regression data, it was observed that those probes where the students' scores correlated highest with the NCE scores on the TerraNova did not necessarily result in significant regression results. Thus, the collinearity statistics were investigated in order to determine whether multicollinearity was affecting the analysis. The results of this initial analysis are presented in Tables 11 and 12. The Variance Inflation Factor (VIF) is typically used to interpret the results of regression in order to understand the proportion of the variance in the independent variable that is related to the other independent variables (O'Brien, 2007). While there are many different numbers used to determine when the VIF has reached unacceptable levels, it is often noted that the ideal VIF should be as close to 1 as possible (O'Brien, 2007). When the VIF reaches higher levels, it is believed that there is an effect of multicollinearity on the analysis. Multicollinearity is defined as an inability to separate the effects of two (or more) variables on an outcome measure (Lynch, 2003). Thus, it was determined that the highly correlated applications assessments affected the calculations of individual predictors, and this was evident from the VIF scores obtained from the analysis which are presented in Tables 11 and 12. The correlations among the applications assessments affect the regression analysis calculations resulting in erroneous data that suggest that variables with lesser correlations are most predictive of performance on the dependent variable.

Table 11. Second Grade Correlation and Collinearity Statistics

		Correlation	Significance	Tolerance	VIF
Computations					
	October	.510	.018	.615	1.627
	January	.321	.867	.539	1.854
	May	.230	.140	.675	1.482
Applications					
	October	.613	.068	.484	2.065
	January	.716	.097	.246	4.061
	May	.749	.000	.291	3.441

Table 12. Third Grade Correlation and Collinearity Statistics

		Correlation	Significance	Tolerance	VIF
Computations					
	October	.496	.170	.611	1.636
	January	.201	.002	.829	1.207
	May	.373	.567	.710	1.409
Applications					
	October	.723	.000	.457	2.188
	January	.730	.101	.319	3.135
	May	.803	.000	.270	3.708

To address this concern, the January applications probes, which had not been found to have significant levels of predictive validity, were removed from the analysis, as is often suggested as a solution to compensate for issues of multicollinearity (Lynch, 2003; O'Brien, 2007). The results of these analyses are found in Tables 13 and 14. The remaining separate applications probes were found to have significant levels of predictive validity as expected from the strong correlations that were obtained. The level of predictive validity on the January Applications probes were then assessed by conducting a linear analysis by removing the May Applications probes data. Again, the separate Applications probes, including January, were found to have significant levels of predictive validity as expected from the strong correlations that were obtained.

Table 13. Linear Regression Statistics for Second Grade CBM Probes and TerraNova Total Math Normal Curve Equivalent (NCE) *

Probe (In order of predictive validity)*	Pearson Correlation	Significance
May Applications	.749	.000
October Applications	.613	.007
October Computations	.510	.016
May Computations	.230	.080
January Computations	.321	.874

*January Applications has been removed to avert multicollinearity concerns.

Table 14. Linear Regression Statistics for Third Grade CBM Probes and TerraNova Total Math Normal Curve Equivalent (NCE) *

Probe (In order of predictive validity)*	Pearson Correlation	Significance
May Applications	.803	.000
October Applications	.723	.000
January Computations	.201	.002
October Computations	.496	.090
May Computations	.373	.676

*January Applications has been removed to avert multicollinearity concerns.

The hypothesis that the corrective education teachers who were involved in the administration of these probes would find the assessment tools more acceptable after the final administration of the probes than before the first was also investigated. The mean scores obtained from the nine participating teachers are listed in Table 15. There was no observable trend in the scores obtained pre-administration to either survey administered post-administration, although the Concepts and Applications probes were rated with a slightly lower mean than the mean rating pre-administration and the mean rating of the Computations probes post-administration.

Table 15. Mean Scores Obtained on the ARP-15 during the Separate Administrations

	Pre-administration	Post-Administration	
		Computations	Concepts and Applications
Mean Rating	4.77	4.82	4.46

A paired sample correlated t-test and a nonparametric correlated Wilcoxon t-test were utilized to better understand the relationship of the scores obtained using the ARP-15. Based on the results of the paired sample correlated t-test, no significant results were yielded indicating no relationship between the pre-administration survey and either post-administration survey, as seen in Table 16.

Table 16. Results of Paired Sample Correlation t-Test between the Separate Administrations

	N	Correlation	Significance
Pre-survey & Computations Post	9	.311	.415
Pre-survey & Applications Post	7	.775	.041

A nonparametric correlated Wilcoxon t-test was used in order to account for differences in two paired variables. Because a different number of individuals were administered the ARP-15 pre-administration of the probes than completed the different version of the post-administration survey, the Wilcoxon t-test was used to account for this difference. However, no significant relationship was found using this statistical analysis, as shown in Table 17.

Table 17. Results of Wilcoxon Signed Ranks Test

	N	Asymp. Significance (2-tailed)
Pre-survey & Computations Post	9	.889
Pre-survey & Applications Post	7	.735

Overall, the following results were obtained. The curriculum-based measurements were found to be reliable measures, correlating moderately to strongly with each other. The computations probes did not consistently correlate with each other, while the applications probes were found to have a strong correlation with each other. When examining the predictive validity of the CBM probes to the NCE of the TerraNova, it was determined that the concepts and applications probes could significantly predict student performance on the appropriate grade level TerraNova Total Math Normal Curve Equivalent. In addition, it was found that the ARP-15 could not reveal significantly more favorable ratings concerning the use of CBM when comparing a survey administered prior to the first administration than after the last administration.

CHAPTER 5

DISCUSSION

Curriculum-based measurement has proven to be a valuable tool in education for many reasons. Curriculum-based measurement procedures have been used to monitor student academic progress, measure the effectiveness of intervention, and even to predict performance on impending high-stakes assessments. Two forms of math curriculum-based measurement, computation fluency and concepts and applications, were used in this study to monitor the progress of students in the second and third grades. The scores obtained from the use of the curriculum-based measurement tools were then compared to the results of a criterion-referenced assessment, the TerraNova, Second Edition, in order to determine whether these CBM tools would be able to adequately predict performance on the TN2.

The results of this study replicated Connell (2006) by demonstrating that these applications assessments had better concurrent validity with the criterion assessments than the computation assessments. The computation assessments had moderate correlations with the criterion assessments and thus moderate concurrent validity. Furthermore, this study extended Connell (2006) by demonstrating the applications assessments had better predicative validity than the computation assessments. No significant predictive validity was identified for the computations probes. That is, the computation probes were not able to significantly indicate whether a student would perform well on a subsequent administration of the TerraNova Total

Math subtests. The concepts and applications probes, which had already been found to have significant levels of concurrent validity with the Louisiana Education Program for the 21st century (LEAP) and the Iowa Test of Basic Skills (ITBS; Connell, 2006), were found to have predictive validity with the TerraNova, Second Edition, Total Math Normal Curve Equivalent. Each of the three administrations of the concepts and applications probes was found to predict a students' performance on the Total Math subtests of the TerraNova, Second Edition.

Overall, it was found that the concepts and applications probes were better predictors of students' progress towards end of year goals. The concepts and applications probes were also more reliable than the computations probes. That is, the scores obtained by the students on each administration of the concepts and applications probes were significantly correlated with each other, while the scores obtained through each administration of the computations probes were not found to have a significant relationship with each other.

The computations probes data were not unique to this study. Helwig, Anderson, and Tindal (2002) indicated that CBM tasks with more emphasis on conceptual understanding may be better predictors of overall math achievement. This is because knowledge of basic concepts is necessary to successfully apply this knowledge to a higher-level question. In addition, because of the overlap that exists between various domains within mathematics, few studies have been able to adequately produce results to verify this belief. Fuchs and colleagues (2007) did find that multiple skill screeners correlated strongly to outcomes of math assessment and thus provided valid information about math competence across first grade. However, Marston (1989) reported few correlations between math computations CBM and criterion-referenced assessments exceeding .60. Only modest support for the use of math computations CBM was found by Good and Jefferson (1998) as well, because of the moderate levels of construct validity that were identified through a meta-analysis. Foegen and colleagues (2007) described a study that compared five different criterion measures with both reading CBM and math CBM. The results of that study demonstrated that the reading CBM correlations were stronger than the math computations CBM with the criterion measures. The present study replicates those findings in that the highest correlation with the TN2 with the math CBM computations probes was a moderate correlation (.510). Thus, the findings of the current study are similar to what has been reported in the extant literature; math computation is at best a moderate predictor of performance on criterion-referenced assessments.

The results of the computation fluency probes obtained in this study were further investigated, and it appeared that there was an inconsistent trend in the mean scores attained by the students in both the second and third grade. The mean score obtained by both the second and third grade students improved from October to January. However, the mean score obtained by both the second and third grade students declined from January to May. Though it is not possible to determine exactly why this change occurred, there is anecdotal information that suggests a possible cause. Direct observations of the computation data sheets indicated that the students attempted to answer more types of computations problems on the May administration. That is, many students skipped around on the first administration of the computation probe searching for known problems. One the second administration, most students attempted the problems in the order in which they were presented (rather than skipping around). As such, the students likely attempted to answer unfamiliar problems. This may be because the students had learned new skills by May, such as subtraction with borrowing, and therefore were attempting to use the skill on this probe when they would have previously skipped the problem. Fortunately, the students were accurate, though they were not yet fluent with these skills. Thus, it appeared that the students sacrificed fluency for accuracy. The computation probes were a measure of fluency and only digits correct per minute were considered in the scoring of these assessments. Therefore, if a student spent a long time answering one question, he/she would have less time to answer more problems on the page and therefore have a lower score. A downward trend was, thus, observed in the mean digit correct per minute scores of the students in both the second and third grades.

Christ and colleagues (2008) summarized the results other studies that examined computations CBM by asserting that the results of computations CBM should only be used with limitations. Christ and colleagues (2008) indicated that computations CBM only assess a student's ability to utilize basic facts and not the applications of these facts to higher level problems. Therefore, it appears that the information gleaned from the computation fluency probes may provide an incomplete profile of a student's progress towards meeting end of year goals.

In 1989, the National Council of Teachers of Mathematics indicated that a mathematics curriculum should focus less on the memorization of basic skills and focus more on the conceptual understanding of mathematical procedures (Jiban & Deno, 2007; McComas, 2000). Few studies have focused on the development of concepts and applications probes, although it has been acknowledged that effective problem solving involves the integration of both computations and applications skills and that more attention should be paid to the application of math skills within the typical math curriculum (Jiban & Deno, 2007; McComas, 2000; Thurber et al., 2002). A number of studies, such as the ones conducted by Connell (2006), Fuchs and colleagues (1994), Helwig and colleagues (2002), Jitendra and colleagues (2005), and Shapiro and colleagues (2006), have all shown that measures of concepts and applications can be strong predictors of performance on state accountability tests in math. Predictive validity has been found through the previous studies between concepts and applications probes and a variety of state assessments including the Pennsylvania System of School Assessment (PSSA), the Minnesota Comprehensive Assessment (MCA), the Louisiana Education Program for the 21st century (LEAP) and the Iowa Test of Basic Skills (Connell, 2006; Jiban & Deno, 2007; Shapiro et al., 2006).

This study successfully demonstrated the predictive validity of measures of concepts and applications to the TerraNova, Second Edition, Total Math Normal Curve Equivalent score. These concepts and applications probes developed by Connell (2006) measured the student's application of math skills to six distinctive areas of math instruction: number sense, data analysis, patterns, algebra, measurement, and geometry. These measures were identified as stronger tools for measuring student progress towards meeting the end of year expectations as measured through the Normal Curve Equivalent of the TerraNova, Second Edition, Total Math subtests. However, when assessing the predictive validity of the individual assessments, issues of multicollinearity were identified. In other words, because the scores obtained from the different administrations of the concepts and applications probes correlated so highly with each other, the data analysis tool identified two different administrations as just one single administration. The correlation was so high that, it appeared to be only one set of data instead of two. To remedy this inconsistency, the data from one administration of the concepts and applications probes were removed and the analysis was completed again. When the analysis was completed again, the different sets of data were identified individually, and each administration of the concepts and applications probes was found to have significant levels of predictive validity.

The concepts and applications CBM probes had the most significant levels of predictive validity of all the administrations of both CBM types, as measured through a linear regression. The May applications probes had the strongest levels of predictive validity and the highest correlations with the TN2 of all the administrations of both forms of CBM in both the second and third grade participants. This is likely because it was administered closest to the administration of the TerraNova, Second Edition. Therefore, the students had received the majority of instruction that would enable them to meet the end-of-year goals.

It was interesting to note that the mean score of the students in both second and third grade improved with each administration. This suggests that these probes may be useful tools for monitoring the progress of students, not just predicting performance on end of year assessments. It is possible that this tool may be sensitive enough to gauge the effect classroom instruction is having on a student’s ability to understand and use mathematical strategies. Concepts and applications probes may be one way to discover whether the students have gained the ability to use basic skills beyond traditional math computation worksheet designs. If a student performs poorly on a CBM probe assessing math concepts and applications skills, the teacher will have the ability to modify instruction and monitor whether the change has influenced the student’s performance. Because all CBM probes, including the concepts and applications probes used in the current study, can be administered multiple times over short intervals, it is possible for teachers to monitor how well the change in instruction has benefitted the student’s ability. The teacher may then be able to determine the best possible way to instruct the students in math.

However, additional research will be necessary to determine if the suggested response to instruction can be demonstrated through empirical analysis.

A second purpose of this study was to investigate the acceptability of these curriculum-based measurements as reported by corrective education teachers. These corrective education teachers had no previous experience with the procedures or the theory behind the use of curriculum-based measurement. Thus, they were given the ARP rating scale two times across the year in order to understand their perceptions of the tools. The first time they were asked to rate the CBM tools was immediately following their first exposure to the procedures. They had been trained and had reviewed the probes, and then were asked to rate the computations and concepts and applications assessments. Then, after the final administration of their first year of use, the corrective education teachers were again asked to rate the CBM tools, this time rating the two types as separate entities on two separate rating scales. The goal was to measure teachers' perceptions of these assessments once they were asked to administer, score and use them for instructional modifications.

The mean ratings reported by the corrective education teachers did not indicate either a positive or negative response to the procedures on any of the three rating scales completed. The scale asked the teachers to rate the procedures on a scale of "one" to "six," with "six" indicating that he or she strongly favored the instrument. The mean scores obtained on all three rating scales ranged from 4.46 to 4.82. It was not possible to indicate strong feelings of like or dislike towards the probes and procedures. No significant difference could be found between the mean scores on the three different administrations of the rating scales. Thus, the corrective education teachers did not reveal any change in their rating of these probes between their first use of the probes and their final use. No significant growth or decline was noted between the ratings on the initial scale and the two final scales.

Wolf (1978) indicated that it is important for an intervention or procedure to be acceptable to the people being asked to use it; otherwise, they are likely to avoid the procedures. Kazdin (1977) indicated that within a school setting, an intervention should only be considered effective if the individuals using the treatment have found it to be acceptable because they are more likely to use it again. Allinder and Oats (1997) found that teachers who rated CBM procedures to be more acceptable had increased fidelity implementing the CBM protocol. In addition, Eckert, Hintze, and Shapiro (1999) noted that perceived acceptability of an assessment tool influenced the use of the assessment tool as well as the effect of the assessment tool. Thus, it was important that the CBM procedures be found to be acceptable assessments by the corrective education teachers in order to ensure that the procedures would be used with fidelity in the future.

While it is not possible to qualitatively understand the mediocre ratings of the CBM probes and procedures, possible explanations emerged to offer some analysis. When the applications probes were developed and implemented by Connell (2006), and the teachers were asked to rate them, the overall ratings were very high (5.8 of 6). However, the teachers were not asked to administer, score, and interpret the results of the CBM probes. The teachers in the current study were asked to complete all these tasks. It is possible that the teachers regarded the CBM procedures not as an important information gathering tool, but as an additional task that must get completed during the school day. Thus, the CBM probes were not initially seen as helpful but were seen to be an additional stressor. It is possible that the amount of work required for the successful administration, scoring and analysis of the probes outweighed the implications of the results.

Another theory that emerged to explain why the computations probes were rated slightly higher than the concepts and applications probes also reflects the corrective education teachers' concerns for time during the school day. The computations probes were quickly administered in either two or three minute intervals. Because of the brief time of administration, the students were only able to answer a limited number of problems. Thus, the corrective education teachers had few problems to score and were able to complete the process in a relatively brief period of time. However, the applications probes were administered during an untimed period and may have taken a period of up to thirty minutes for the students to complete. Because of the necessary component of reading the questions on these probes, it is likely that the students needed a great deal of interaction with the teacher during the administration, both to stay on task during the assessment and for assistance understanding the problems. In addition, the applications probes were longer and required more time to be scored. Overall, it is possible that the computations probes were slightly more acceptable than the concepts and applications probes because of the disparity in the length of time necessary to both administer and score the different types of probes used.

Limitations

The results of this study appear to provide new and interesting information to the study of curriculum-based measurement. However, there are limitations to the evaluation that need to be considered. One of the primary limitations to this study was that only one set of computations probes was used at each administration. Hintze, Christ, and Keller (2002) determined that the median score of three administrations should be used when scoring CBM math multi-skill computations probes. While only one probe can be used for single-skill probes, it is ideal to take the median score of three probes to assess a student's performance on multi-skill computations probes. However, this procedure was not followed in the current study. It is unclear how this impacted the overall results of the study, but it is suggested that three multi-skill probes should be used at each administration in any future replications of this study.

Second, in addition to not using the median of the three multiple skill computation probes, all three multiple skills assessments were different (though alternate) forms. Though the same skills (e.g., subtracting with borrowing, addition with regrouping) were assessed on each form, the algorithms for each of the probes changed as the problems changed. It is possible that the different problems may have negatively affected the outcome data for the computation assessments. In addition, it may have been invalid to compare the results of the three alternative versions to the results of the same version concepts and applications probes.

The administrators of the math CBM computation assessments were each provided with a script in order to insure a high degree of treatment integrity. However, no explicit procedural fidelity checks were put in place during the year. Due to the large number of individuals who administered the probes in various settings, it is possible that the administration was inconsistent across settings or even across time periods. This may have affected the performance of the students in a negative way. It is not possible to determine the effect that this may have had on the overall results of this study.

Because of the high levels of correlations when comparing the results of the multiple administrations of concepts and applications probes, an issue of multicollinearity arose. This affected the assessment of predictive validity by masking the scores from one administration and making it appear to be a part of another administration. Because of this, only two of the three highest correlating administrations of curriculum-based measurement were identified as significantly predicting performance on the TerraNova Total Math subtests. However, when the various administrations were segregated from each other, each of the administrations of the concepts and applications probes were found to be able to predict performance on the end-of-year assessment. Thus, it does appear that the high correlations obtained between the various administrations of the concepts and applications probes may have been a limitation to the analysis and interpretation of the results of this study.

Because of the specialized population of participants in this study, there may be limitations to how far the results of the study can be generalized. The participants of the study were all students in nonpublic schools in an urban environment. While there was a great deal of variety in ethnicity and socio-economic status, it should be noted that the students do have at least one characteristic in common. They have sought out an education setting alternative to the local public school. The way that these students are educated may be different than in other schools because of different missions of the schools than typical public schools. In addition, these schools tend to use the results from standardized assessments in a different way than most public schools. Typically, the results of standardized assessments in nonpublic schools are used as information gathering tools, as opposed to the decision making tools they are often used as in public schools. The funding for nonpublic schools is typically not dependent upon the results of these assessments, and thus they are not considered to "high-stakes tests," as they are in public schools.

Another limitation of this study and most studies involving curriculum-based measurement is the amount of time necessary to create, administer, and score the probes in order to obtain the information. Because curriculum-based measurement is often seen as something additional to the typical daily requirements of a teacher, time is usually not allotted for the teachers to accomplish these tasks. Thus, it is difficult for teachers to regard these procedures with high levels of acceptability. However, it is important that these procedures are regarded with high levels of acceptability because that leads to higher levels of integrity (Allinder & Oats, 1997; Wolf, 1977). The more the procedures are found to be acceptable, the more likely it is that the teachers will use them as they are intended.

Future directions

Future replications of this study should use a more heterogenous population of participants so that the results may be generalized to as many different classroom environments as possible. In addition, future research should use three different computations probes at each administration and the median score should be regarded as the measure of a student's performance on the fluency assessment.

It has been noted that in order for curriculum-based measurement to be regarded as having diagnostic efficiency, there are four statistical measures that must be completed: sensitivity, specificity, positive predictive power, and negative predictive power (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006). It does appear that the concepts and applications probes used in the current study have met one of these four measures, sensitivity. That is, because of the high levels of statistical predictive validity, there is high probability that when a student performs poorly on the probe, it is highly likely the student will perform poorly on the standardized achievement test. However, future studies should examine explicitly how well these concepts and applications probes meet the remaining expectations for diagnostic accuracy: specificity, positive predictive power, and negative predictive power (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro et al., 2006).

Another possibility for future research would be a measure of sensitivity to the effectiveness of an intervention. A study by Fuchs and colleagues (1994) used CBM to measure how well consultation practices improved the math achievement of students with a variety of educational classifications. Another study by Fuchs and colleagues (2007) suggested that multi-skill math computations CBM may be used to determine a student's response to an intervention within a second or third tier of a Response to Intervention framework. Because it appeared that the students in this study progressively improved the mean score from each administration to the next, there is a suggestion that the concepts and applications probes may be sensitive to instruction. That is, the students performed better as they were taught more in the classroom. Thus, the sensitivity of these probes should be measured in relation to specific interventions.

In addition, future research should explore the use of concepts and applications probes with different age groups. The concepts and applications probes created by Connell (2006) are available for more grade levels than second and third grade. Therefore, this study should be replicated with different grade levels in order to further the predictive validity of these probes to the TerraNova, Second Edition, Total Math Normal Curve Equivalent.

Conclusion

The field of education is always changing. With developments in education research and alterations of education law, it is important that the practices of those within the field adapt to meet the newly identified needs. Ultimately, the purpose of the field of education is to benefit the future of students by adequately preparing them in the present day. It is the responsibility of the people working in the field of education to find, develop, and use the best possible assessment and instructional tools to ensure that the ultimate purpose of education is met.

Recent federal legislation has changed the way a school psychologist performs his or her job. According to the latest version of the special education law, the IQ/Achievement discrepancy model is no longer the only way to identify a student for special education services. The Response to Intervention process can be used as an alternative, but in order for this process to be effective, data must be kept to be sure how effective an intervention has been (Ardoin, 2006; Barnett, et al., 2007; Daly et al., 2007; Fuchs & Fuchs, 2007; Fuchs et al., 2008; Hilton, 2007; Shinn, 2007; Willis & Dumont, 2006; Zirkel, 2007).

Additional federal legislation has mandated that a school prove its effectiveness based on that school's ability to obtain adequate yearly progress, as measured through statewide accountability testing (Katsiyannis et al., 2007; NCLB, 2001; Silberglitt et al., 2006). The No Child Left Behind legislation (NCLB, 2001) has added pressure to schools to ensure that the students are able to perform at the highest level possible on these state-mandated assessments. Thus, schools have found it helpful to assess the students' progress towards meeting end of year goals that will be included on the state-mandated assessment. By obtaining information about the students' progress in specific skills, it is possible to alter instruction, and if necessary, to ensure that the students are able to perform at the highest level possible on the assessment.

The needs of schools that have recently been identified as a result of these changes in federal legislation can easily be met by one particular set of procedures. Curriculum-based measurement, or CBM, is widely used within the field of education because it has been found to be a fast, reliable, and valid measure of student progress basic skill content areas (Cusumano, 2007; Deno, 1989; Deno, 2003; Fewster & MacMillan, 2002; Fuchs & Fuchs, 2002; Fuchs et al., 1994; Hosp et al., 2007; Marston, 1989; Stecker et al., 2005). CBM tools have been used to monitor a student's progress with or without an intervention in place, as well as predict a student's performance on a standardized assessment (Connell, 2006; Foegen & Deno, 2001; Foegen, Jiban, & Deno, 2007; Francis et al., 2005; Fuchs & Fuchs, 2002; Fuchs & Fuchs, 2007; Helwig et al., 2002; Hintze & Silberglitt, 2002; Hosp & Hosp, 2003; Jiban & Deno, 2007; Jitendra et al., 2005; McCardle et al., 2008; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Silberglitt & Hintze, 2005; Stecker et al., 2005). The information obtained from CBM tools allows educators to change instruction in enough time to increase student knowledge and eventually performance on the state-mandated tests (Helwig et al., 2002; Jiban & Deno, 2007). When student performance improves, the accountability of the school also improves (Helwig et al., 2002; Hintze & Silberglitt, 2005; Jiban & Deno, 2007; Silberglitt et al., 2006; Silberglitt & Hintze, 2005).

CBM is a sensitive measure of basic skill progress over time; moreover, it has been used as a dynamic indicator of overall academic growth and development (Hasbrouck et al., 1999; Hintze, Christ, & Keller, 2002; Hosp et al., 2007; Hintze & Silberglitt, 2005; Stecker et al., 2005). However, it is important that all of the basic skills that are being introduced to the students can be assessed through the use of CBM. The concepts and applications probes used in this study have ensured that a student's ability to apply math skills to more "real life" situations can be assessed at short periods of time over brief intervals. While this assessment is important to understand the progress of the individual students towards meeting end-of-year goals, these probes may also provide insight into the student's overall math ability. Because these probes have been found to predict a student's performance on an overall measure of math achievement, these probes may also be helpful in determining whether alterations need to be made to instruction in order to ensure a school can be found to be accountable.

Overall, the results of this study can be applied to the practice of education in a number of ways. By using these probes, which have been identified to have significant levels of reliability, concurrent validity, and predictive validity, a student's performance in math concepts and applications can be positively affected. It will be possible to alter instruction early enough in the academic year to ensure that the students have the opportunity to learn as much as possible in the appropriate grade level in the area of concepts and applications. These probes can ensure that the students have acquired the basic knowledge of concepts and applications at an early grade level to provide a strong foundation of math skills for the future academic experiences of the individual students. It appears that these concepts and applications probes have the ability to affect both the short-term and long-term future of math education. Thus, these concepts and applications probes will ensure that the people working in the field of education will be better able to uphold their responsibility to ensure that the ultimate purpose of education is met: students are adequately prepared to handle the academic and civic demands of the future.

REFERENCES CITED

Agresta, J. (2004). Professional role perceptions of school social workers, psychologists, and counselors. *Children & Schools, 26*(3), 151-163.

Allinder, R. M. (1995). An examination of the relationship between teacher efficacy and curriculum-based measurement and student achievement. *Remedial and Special Education, 16*(4), 247.

Allinder, R. M., & BeckBest, M. A. (1995). Differential effects of two approaches to supporting teachers' use of curriculum-based measurement. *School Psychology Review, 24*(2), 287.

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*(2), 113.

Ardoin, S. P. (2006). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*(6), 713.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23,* 362-380.

Barnett, D. W., VanDerHeyden, A. M., & Witt, J. C. (2007). Achieving science-based practice through Response to Intervention: What it might look like in preschools. *Journal of Educational and Psychological Consultation*, *17*, 31-54.

Braden, J. P. (2007). Using data from high-stakes testing in program planning and evaluation. *Journal of Applied School Psychology*, *23*(2), 129-150.

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24*(4), 235-245.

Chafouleas, S. M., Riley-Tillman, T. C., & Eckert, T. L. (2003). A comparison of school psychologists' acceptability, training, and use of norm-referenced, curriculum-based, and brief experimental analysis methods to assess reading. *School Psychology Review, 32*(2), 272-281.

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

Cizek, G. J. (2005). Review of TerraNova, The Second Edition. In B. S. Plake, L. L. Murphy, & R. A. Spies (Eds.) *The sixteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Connell, J. E. J. (2006). Constructing a math applications, curriculum-based assessment: An analysis of the relationship between applications problems, computation problems and criterion-referenced assessments. *Dissertation Abstracts International*, *66* (7-B), 3393.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, *7*(4), 303-323.

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

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

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

Eckert, T. L., Hintze, J. M., & Shapiro, E. S., (1997). School psychologists' acceptability of behavioral and traditional assessment procedures for externalizing problem behaviors. *School Psychology Quarterly*, *12*(2), 150 – 169.

Eckert, T. L., Hintze, J. M., & Shapiro, E. S., (1999). Development and refinement of a measure for assessing the acceptability of assessment methods: The assessment rating profile – revised. *Canadian Journal of School Psychology*, *15*, 21 – 42.

Eckert, T. L., Shapiro, E. S., & Lutz, J. G. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review, 24*(3), 497.

Erchul, W. P. & Martens, B. K. (2002). *School consultation: Conceptual and empirical bases of practice*. New York, NY: Springer Science+Business Media, Inc.

Fewster, S., & MacMillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149.

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools, 43*(8), 835-853.

Foegen, A. & Deno, S. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education, 35*(1), 4-16.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring in mathematics: A review of the literature. *The Journal of Special Education, 41*(2), 121-139.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*(2), 98-108.

Fuchs, L. S. (1992). Classwide decision-making with computerized curriculum-based measurement. *Preventing School Failure, 36*(2), 30-33.

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

Fuchs, L. S., & Fuchs, D. (2001). Principles for sustaining research-based practice in the schools: A case study. *Focus on Exceptional Children, 33*(6), 1.

Fuchs, L. S. & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education*, *77*(2), 64-84.

Fuchs, L. S. & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, *39*, 14-20.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children*, *73*(3), 311-330.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*, *60*(6), 518-537.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, *28*(3), 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubek, P., & Stecker, P. M. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique*, *19*, 23-49.

Fuchs, L. S., Fuchs, D., & Zumeta, R. O. (2008). Response to intervention: A strategy for the prevention and identification of learning disabilities. In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 115-135). New York, NY: Springer Publishing Company, LLC.

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

Good, R. H. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York, NY: The Guilford Press.

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating Response to Intervention and cognitive assessment methods. *Psychology in the Schools*, *43*(7), 753-770.

Harry, B. & Klingner, J. (2007). Discarding the deficit model. *Educational Leadership*, *64*(5), 16-21.

Hasbrouck, J. E., Woldbeck, T., Ihnot, C., & Parker, R. I. (1999). One teacher's use of curriculum-based measurement: A changed opinion. *Learning Disabilities Research & Practice*, *14*(2), 118-126.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education*, *36*(2), 102-112.

Hilton, A. (2007). Response to intervention: Changing how we do business. *Leadership, 36*(4), 16.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*(4), 514-528.

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

Hogan, T. P. (2003). *Psychological testing: A practical introduction*. New York, NY: John Wiley & Sons, Inc.

Hosp, M. K., & Hosp, J. L. (2003). Curriculum-based measurement for reading, spelling, and math: How to do it and why. *Preventing School Failure, 48*(1), 10-16.

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

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, §614, 118 Stat. 2647 (2004).

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.

Jitendra, A. K., Sczesniak, E., & Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third graders. *School Psychology Review*, *34*(3), 358-371.

Johnson, R. L. & Mazzie, D. (2005). Review of TerraNova, The Second Edition. In B. S. Plake, L. L. Murphy, & R. A. Spies (Eds.) *The sixteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, *23*(2), 65-86.

Katsiyannis, A., Zhang, D., Ryan, J. B., & Jones, J. (2007). High-stakes testing and students with disabilities. *Journal of Disability Policy Studies*, *18*(3), 160-167.

Kaufman, A. S. (2000). Intelligence tests and school psychology: Predicting the future by studying the past. *Psychology in the Schools*, *37*(1), 7-16.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification, 1*(4), 427-452.

Kazdin, A. E., French, N. H., & Sherick, R. B. (1981). Acceptability of alternative treatments for children: Evaluations by inpatient children, parents, and staff. *Journal of Consulting and Clinical Psychology, 49*(6), 900-907.

Lichtenstein, R. & Fischetti, B. A. (1998). How long does a psychoeducational evaluation take? An urban Connecticut study. *Professional Psychology: Research and Practice*, *29*(2), 144-148.

Lynch, S. M. (2003, February). Multicollinearity. Retrieved March 26, 2009, from http://www.princeton.edu/~slynch/SOC_504/multicollinearity.pdf.

Madaus, G. F. & Kellaghan, T. (1993). Testing as a mechanism of public policy: A brief history and description. *Measurement and Evaluation in Counseling and Development*, *26*, 6-10.

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

McCardle, P., Keller-Allen, C., & Shuy, T. (2008). Learning disability identification: How does it apply to English language learners? In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 137-164). New York, NY: Springer Publishing Company, LLC.

McComas, K. K. (2000). Felix Klein and the NCTM's Standards: A mathematician considers mathematics education. *Mathematics Teacher*, *93*(8), 714-717.

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

Naglieri, J. A. & Kaufman, A. S. (2008). IDEIA and specific learning disabilities: What role does intelligence play? In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 165-195). New York, NY: Springer Publishing Company, LLC.

National Association of State Directors of Special Education, Inc. (2005). *Response to intervention: Policy considerations and implementations.* Alexandria, VA: Author.

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

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*, 673-690.

Peterson, K. M. H. & Shinn, M. R. (2002). Severe discrepancy models: Which best explains school identification practices for learning disabilities? *School Psychology Review*, *31*(4), 459-476.

Sattler, J. (2001). *Assessment of children: Cognitive applications, fourth edition*. La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention, third edition*. New York, NY: The Guilford Press.

Shapiro, E. S., Angello, L. M., & Eckert, T. L. (2004). Has curriculum-based assessment become a staple of school psychology practice? An update and extension of knowledge, use, and attitudes from 1990 to 2000. *School Psychology Review*, *33*(2), 249-257.

Shapiro, E. S., & Eckert, T. L. (1992, August). *Acceptability of curriculum-based assessment by school psychologists*. Paper presented at the Annual Convention of the American Psychological Association, Washington, D.C.

Shapiro, E. S., & Eckert, T. L. (1994). Acceptability of curriculum-based assessment among school psychologists. *Journal of School Psychology*, *32*, 167-183.

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

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within Response to Intervention: Research on educational need and benefit from academic intervention. *School Psychology Review*, *36*, 601-617.

Shriberg, D. & Kruger, L. J. (2007). Introduction and overview: High-stakes testing. *Journal of Applied School Psychology*, *23*(2), 1-6.

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

Silberglitt, B. & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment*, *23*, 304-325.

Spear-Swerling, L. (2008). Response to intervention and teacher preparation. In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 273-293). New York, NY: Springer Publishing Company, LLC.

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

Stevens, J. J. & Zvoch, K. (2007). Confirmatory factor analysis of the TerraNova Comprehensive Tests of Basic Skills/5. *Educational and Psychological Measurement*, *67*(6), 976-989.

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

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of Response to Intervention (RtI) model on identification of children from special education. *Journal of School Psychology*, *45*, 225-256.

Willis, J. O., & Dumont, R. (2006). And never the twain shall meet: Can response to intervention and cognitive assessment be reconciled? *Psychology in the Schools*, *43*(8), 901.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools*, *20*, 510-517.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*(2), 203-214.

Zirkel, P. A. (2007). What does the law say? *Teaching Exceptional Children*, *39*, 65-67.

APPENDICES

APPENDIX A

SCRIPT FOR SECOND GRADE COMPUTATIONS CBM

Materials needed for giving CBM math probes

- Student copy of CBM math probe
- Stopwatch
- Pencils for students

Directions for Administration of **Second Grade** Probe

Words written in *italics* should be spoken by the administrating teacher (remedial or classroom).

Distribute the probes to each student.

The sheets on your desk are math facts. As soon as you have the paper, write your first name, your last name, and the date. After you've written your name and the date on the test, put your pencil down so I know you are ready.

There are several types of problems on the sheet. Some are addition and some are subtraction. Look at each problem carefully before you answer it.

When I say "Start," turn them over and begin answering the problems. Start the first problem on the left on the top row [point]. *Work across and then go to the next row. If you can't answer the problem make an "X" on the number of the problem and go to the next one. If you finish one side, go to the back. Are there any questions?*

Say, "*Start*." The examiner starts the stopwatch.

While the students are completing worksheets, the examiner and any other adults assisting in the assessment circulate around the room to ensure that students are working on the correct sheet, that they are completing problems in the correct order (rather than picking out only the easy items), and that they have pencils, etc.

After **2 minutes**, the examiner says, "*Stop*." CBM math probes are then collected for scoring.

APPENDIX B

SCRIPT FOR THIRD GRADE COMPUTATIONS CBM

Materials needed for giving CBM math probes

- Student copy of CBM math probe
- Stopwatch
- Pencils for students

Directions for Administration of **Third Grade** Probe

Words written in *italics* should be spoken by the administrating teacher (remedial or classroom).

Distribute the probes to each student.

The sheets on your desk are math facts. As soon as you have the paper, write your first name, your last name, and the date. After you've written your name and the date on the test, put your pencil down so I know you are ready.

There are several types of problems on the sheet. Some are addition, some are subtraction, some are multiplication, and some are division. Look at each problem carefully before you answer it.

When I say "Start," turn them over and begin answering the problems. Start the first problem on the left on the top row [point]. *Work across and then go to the next row. If you can't answer the problem make an "X" on the number of the problem and go to the next one. If you finish one side, go to the back. Are there any questions?*

Say, "*Start*." The examiner starts the stopwatch.

While the students are completing worksheets, the examiner and any other adults assisting in the assessment circulate around the room to ensure that students are working on the correct sheet, that they are completing problems in the correct order (rather than picking out only the easy items), and that they have pencils, etc.

After **3 minutes**, the examiner says, "*Stop*." CBM math probes are then collected for scoring.

APPENDIX C

SCRIPT FOR CONCEPTS AND APPLICATIONS CBM

<u>Materials needed for giving CBM (**APPLICATION**) math probes</u>

- Student copy of CBM math probe
- Pencils for students
- Each student needs an inch ruler

<u>Directions for Administration of **Application** Probe</u>

Words written in *italics* should be spoken by the administrating teacher (remedial or classroom).

Distribute the probes to each student.

The sheets on your desk are math worksheets. As soon as you have the paper, write your first name, your last name, and the date. After you've written your name and the date on the test, put your pencil down so I know you are ready.

There are several types of problems on the sheet. Read each problem <u>carefully</u> before you answer it.

When I say "Start," turn them over and begin answering the problems. Start at the first problem on the page [point], *and work your way to the last problem without skipping any. If you do not know the answer to a problem, take your best guess and go to the next one. You will not be penalized for putting down the wrong answer, so do not skip any questions. Put your paper in the air when you are finished. Are there any questions?*

Say, "*Start*."

Please note: There is <u>no timing</u> aspect to this assessment.

While the students are completing worksheets, the examiner and any other adults assisting in the assessment circulate around the room to ensure that students are working on the correct sheet, that they are completing problems in the correct order (rather than picking out only the easy items), and that they have pencils, etc.

When the students have completed the assessment and are holding it in the air, collect the CBM math applications probes for scoring.

APPENDIX D

SECOND GRADE (OCTOBER) COMPUTATIONS CBM

Curriculum-Based Assessment Mathematics
Multiple-Skills Computation Probe: Student Copy (October)

Student: _________________ Date: ___________________

1. 30 + 7	2. 2 + 1	3. 83 - 21	4. 60 - 33
5. 7 + 8	6. 2 + 56	7. 7 + 21	8. 63 - 12
9. 81 - 63	10. 7 + 5	11. 1 + 8	12. 31 + 6
13. 45 - 12	14. 84 - 17	15. 4 + 7	16. 2 + 6

17. 2 + 6	18. 83 - 41	19. 66 - 48	20. 6 + 2
21. 34 + 1	22. 5 + 76	23. 27 - 15	24. 93 - 75
25. 8 + 2	26. 84 + 5	27. 7 + 61	28. 27 - 11
29. 31 - 12	30. 8 + 6	31. 44 + 4	32. 8 + 12
33. 33 - 11	34. 84 - 38	35. 9 + 2	36. 2 + 31
37. 8 + 7	38. 55 - 42	39. 53 - 46	40. 5 + 2

APPENDIX E

SECOND GRADE (JANUARY) COMPUTATIONS CBM

Curriculum-Based Assessment Mathematics
Multiple-Skills Computation Probe: Student Copy (January)

Student: ________________ Date: __________________

1. 42 +82	2. 91 -52	3. 1 + 8	4. 36 -25
5. 2 + 4	6. 2 +15	7. 32 -13	8. 1 + 1
9. 76 -62	10. 6 + 3	11. 8 + 2	12. 80 -26
13. 7 + 1	14. 76 -66	15. 12 + 5	16. 8 + 7
17. 82 -18	18. 2 + 7	19. 53 -21	20. 86 + 2

21. 1 +30	22. 45 -26	23. 4 + 6	24. 46 -21
25. 36 + 2	26. 7 + 7	27. 90 -85	28. 7 + 5
29. 86 -23	30. 1 + 4	31. 3 + 1	32. 37 -28
33. 8 + 7	34. 36 -25	35. 84 + 4	36. 22 + 3
37. 24 -15	38. 5 + 6	39. 35 -13	40. 12 +66

APPENDIX F

SECOND GRADE (MAY) COMPUTATIONS CBM

Curriculum-Based Assessment Mathematics
Multiple-Skills Computation Probe: Student Copy (May)

Student: __________________ Date: ___________________

1. $\begin{array}{r} 2 \\ +56 \\ \hline \end{array}$	2. $\begin{array}{r} 81 \\ -47 \\ \hline \end{array}$	3. $\begin{array}{r} 4 \\ +\ 5 \\ \hline \end{array}$	4. $\begin{array}{r} 27 \\ -13 \\ \hline \end{array}$
5. $\begin{array}{r} 6 \\ +\ 8 \\ \hline \end{array}$	6. $\begin{array}{r} 71 \\ +\ 1 \\ \hline \end{array}$	7. $\begin{array}{r} 52 \\ -13 \\ \hline \end{array}$	8. $\begin{array}{r} 7 \\ +\ 1 \\ \hline \end{array}$
9. $\begin{array}{r} 82 \\ -51 \\ \hline \end{array}$	10. $\begin{array}{r} 6 \\ +15 \\ \hline \end{array}$	11. $\begin{array}{r} 6 \\ +\ 2 \\ \hline \end{array}$	12. $\begin{array}{r} 83 \\ -14 \\ \hline \end{array}$
13. $\begin{array}{r} 6 \\ +\ 5 \\ \hline \end{array}$	14. $\begin{array}{r} 33 \\ -23 \\ \hline \end{array}$	15. $\begin{array}{r} 75 \\ +35 \\ \hline \end{array}$	16. $\begin{array}{r} 2 \\ +\ 7 \\ \hline \end{array}$
17. $\begin{array}{r} 23 \\ -18 \\ \hline \end{array}$	18. $\begin{array}{r} 2 \\ +\ 6 \\ \hline \end{array}$	19. $\begin{array}{r} 35 \\ -21 \\ \hline \end{array}$	20. $\begin{array}{r} 34 \\ +\ 4 \\ \hline \end{array}$

21. 1 + 2	22. 63 -17	23. 4 + 5	24. 43 -11
25. 8 + 4	26. 76 + 2	27. 52 -43	28. 7 + 2
29. 58 -12	30. 61 + 5	31. 45 + 4	32. 33 -16
33. 8 + 4	34. 64 -31	35. 5 + 4	36. 14 + 5
37. 96 -37	38. 1 + 1	39. 34 -23	40. 4 +93

APPENDIX G

SECOND GRADE CONCEPTS AND APPLICATIONS CBM

Second Grade
Multi-Assessment Math Probe

Teacher_______________________
School________________________
Student_______________________
Date__________________________

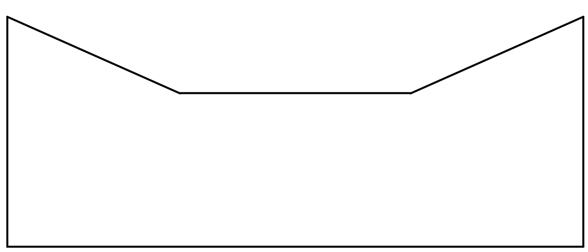
N1. 11 – 5 = _____

A1. Pick > , < , or =

45 ☐ 54

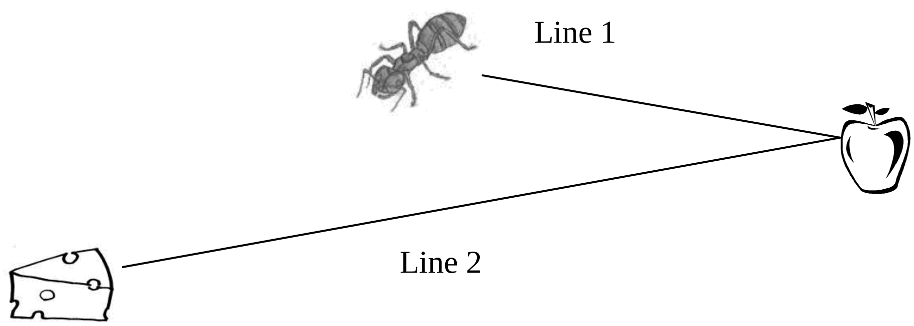
a) >
b) <
c) =

G1. Write how many sides.



_____ sides

M1. Use an inch ruler to measure.
Write the answer.

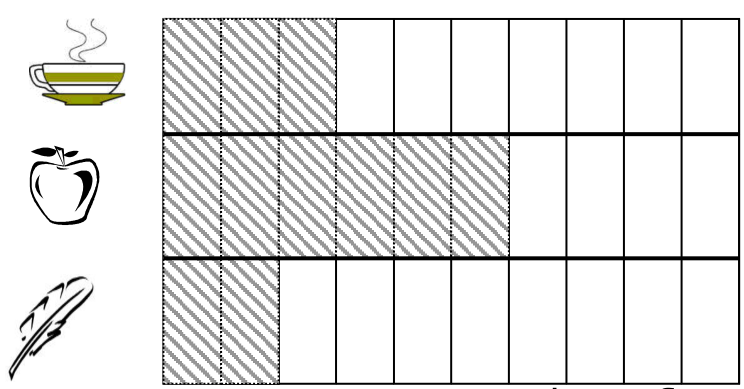


1. Line 1 is about _____ inches.
2. Line 2 is about _____ inches.

P1. Write the missing number:

3, _____, 5, 6

D1. Use the graph.



How many more cups than feathers? _____

N2.

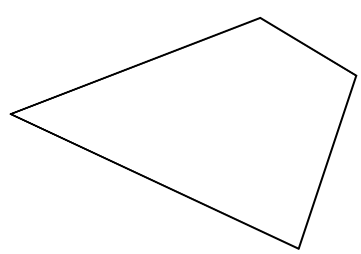
```
    8
    3
+   1
-----
```

A2. Pick > , < , or =

74 ☐ 77

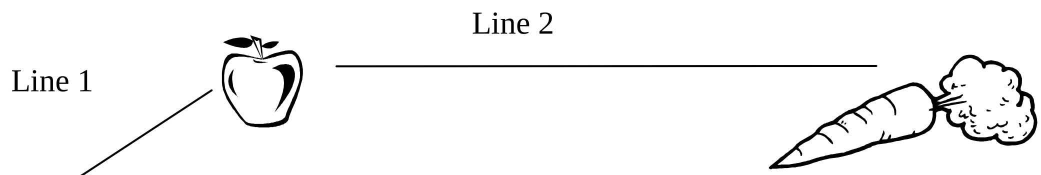
a) <
b) >
c) =

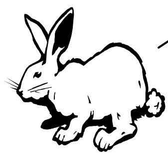
G2. Write how many corners.



_______ corners

M2. Use an inch ruler to measure.
Write the answer.





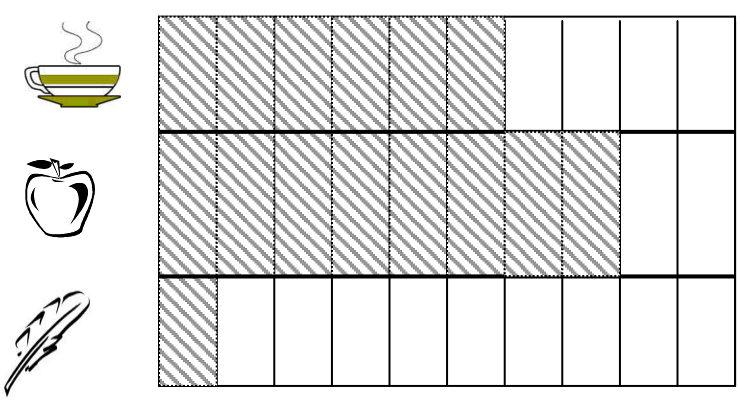
1. Line 1 is about _____ inches

2. Line 2 is about _____ inches

P2. What the missing number.

8, _____, 6

D2. Use the graph.



How many fewer feathers than apples? _____

N3. Write the number that completes the number sentence.

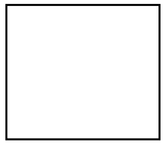
4 + ______ = 11

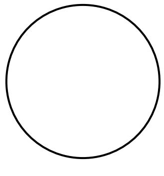
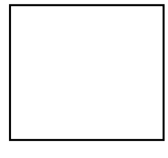
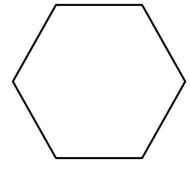
A3. Pick > , < , or =

56 ☐ 52

a) <
b) >
c) =

G3. Circle the one that is the same.

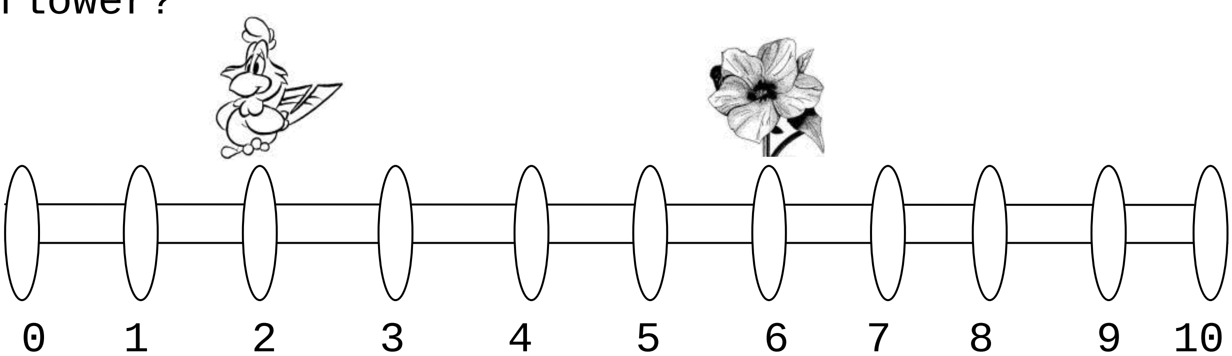


a)  b)  c) 

M3. Circle the best estimate.

a) more than a pound
b) about a pound
c) less than a pound

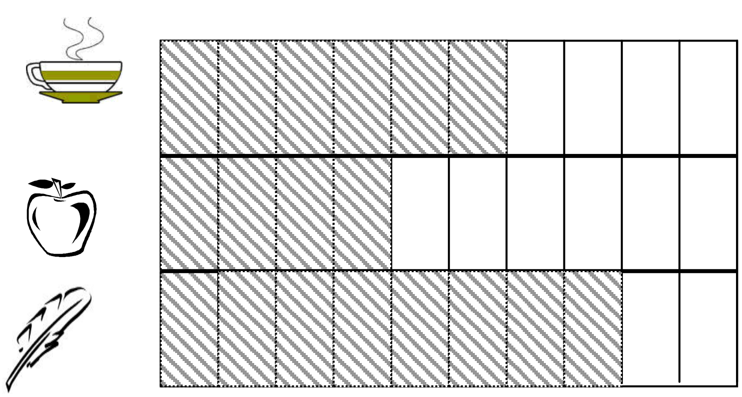
P3. What numbers are between the bird and the flower?



What numbers are between the bird and the flower?

_____, _____, _____.

D3. Use the graph.

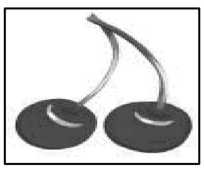
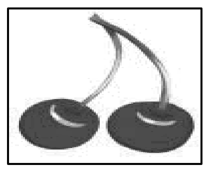
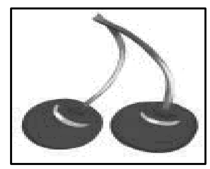


How many more cups than apples? _____

N4. On the field day 100 children had pickles. During the day 63 children ate their pickle. How many children still have pickles?

_____ children

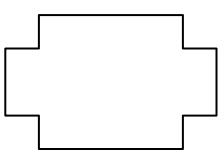
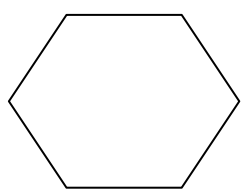
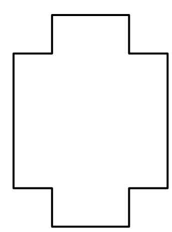
A4. Write how many.

________ + ________ + ________ = ________

________ sets of ________ = ________

G4. Circle the one that is different.

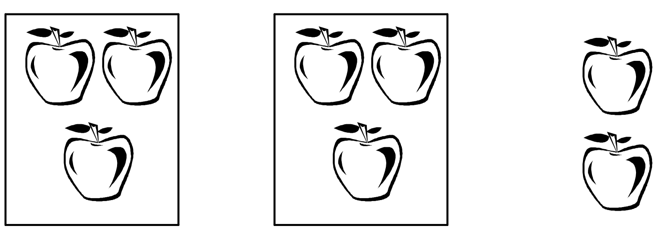
M4. Circle the best estimate.

a) more than a gallon
b) about a gallon
c) less than a gallon

P4. Write the missing number.

4, 8, 12, ______, 20.

D4. Find how many.



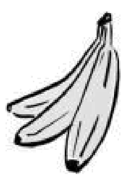
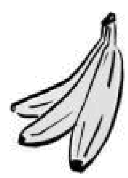
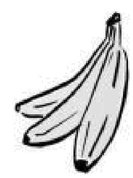
-How many apples?

a)6
b)8
c)10

N5. There are 5 plates on the table. On each plate are 3 grapes. How many grapes are there?

a) 15 grapes
b) 8 grapes
c) 2 grapes

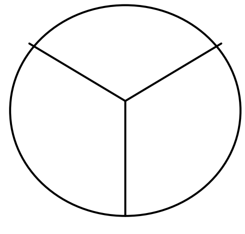
A5. Write how many.

________ + ________ + ________ = ________

________ sets of ________ = ________

G5. Color the figure to show the fraction.
-two thirds



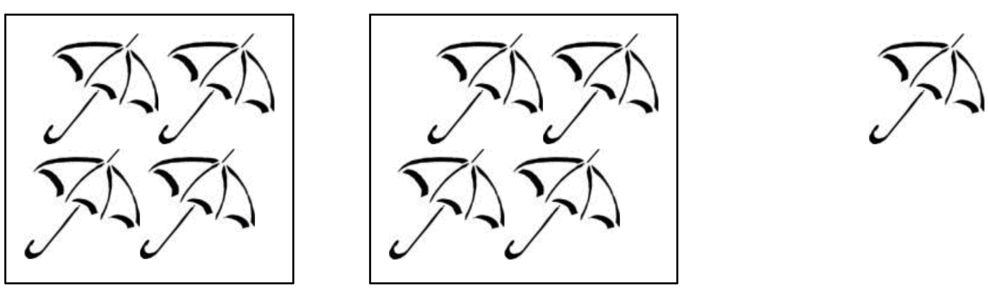
M5. Circle the best estimate.

a) more than a liter
b) about a liter
c) less than a liter

P5. Write the missing number.

35, _____, 45, 50

D5. Find how many.



-How many sets? __________

N6. Tanisha has $1.50 to spend at the movies. April has $2.00 to spend. How much do Tanisha and April have to spend altogether?

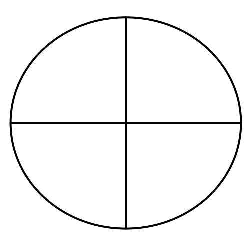
a) $2.50
b) $3.50
c) $4.50

A6. Write how many.

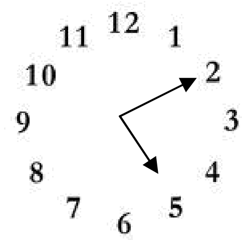
________ + ________ = ________

________ sets of ________ = ________

G6. Color the figure to show the fraction.
-one fourth



M6. Find the time.

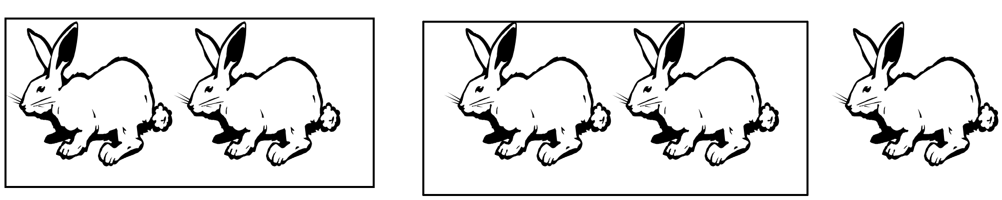


a) 5:50
b) 5:10
c) 6:10

P6. Write the missing number.

______, 50, 60, 70.

D6. Find how many.



-How many left over?

a) 0
b) 1
c) 2

APPENDIX H

THIRD GRADE CONCEPTS AND APPLICATIONS CBM

Third Grade
Multi-Assessment Math Probe

Teacher________________________
School_________________________
Student________________________
Date___________________________

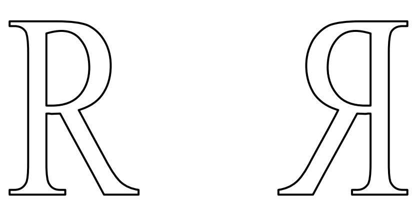
N1. Write the number in standard form.

500 + 60 + 3

A1. Complete the number sentence.

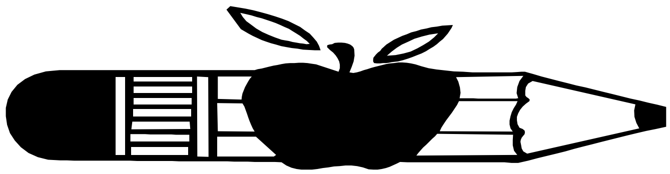
5 + _______ = 12

G1. Choose slide, flip, or half-turn for the set of figures.



a) slide
b) flip
c) half-turn

M1. Measure. Use a ruler to measure to the nearest inch.



Nearest half inch? _______

P1. Write the product that the symbol stands for.

	1	**2**	**3**	**4**
1	*1*	*2*	*3*	*4*
2	*2*	*	*6*	
3	*3*	*6*	*9*	*12*
4		*8*		*16*

* = ______

D1. Use the chart.
Favorite Class Pets.

Pet	Tally
Dog	卌 |
Cat	|||
Fish	||

How many more students picked dogs rather than cats as their favorite pets? _______

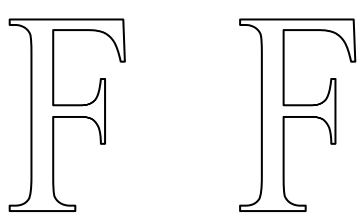
N2. Find the sum.

```
  436
+  77
_______
```

A2. Complete the number sentence.

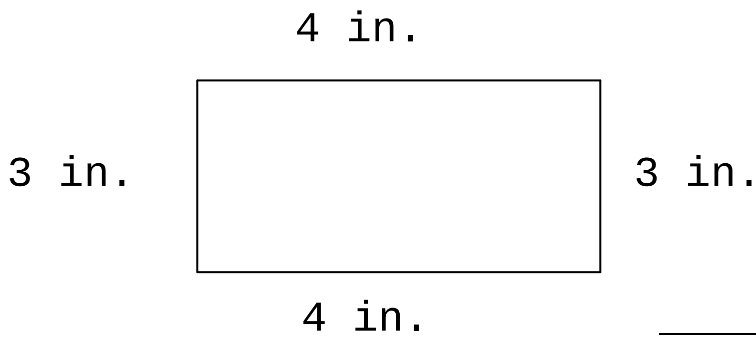
________ + 8 = 17

G2. Choose slide, flip, or half-turn for the set of figures.



a) slide
b) flip
c) half-turn

M2. Write the perimeter.



____

P2. Write the product that the symbol stands for.

	1	**2**	**3**	**4**
1	*1*	*2*		*4*
2	*2*	*4*		*
3			*9*	*12*
4	*4*		*12*	

* = ________

D2. Use the chart.

How many students took part in the survey?

____

Favorite Color

Color	Tally
Red	~~||||~~ |
Blue	||
Black	~~||||~~ |||

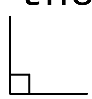
N3. Find the difference.

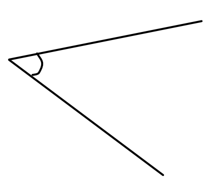
$$\begin{array}{r} 335 \\ -\ 56 \\ \hline \end{array}$$

A3. Write the number sentence that is missing from the fact family.

7 + 2 = 9
2 + 7 = 9
9 - 7 = 2

____

G3. Choose if the angle is <, >, or =
a right angle 



a) <
b) >
c) =

M3. Write the correct unit to measure the following.

-water in a bathtub

a) ounce
b) cup
c) gallon

P3. Write the product that the symbol stands for.

	1	**2**	**3**	**4**
1	*1*			*4*
2		*4*	*6*	
3	*3*		*9*	*12*
4		*8*	*	*16*

* = ________

D3. Use the pictograph.

Money Raised by Bake Sales

How much more money was raised in May than March?

March	$$$$
April	$$
May	$$$$$
Key: $=5 Dollars	

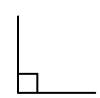
N4. You plan to use 25 lbs. of pork and 38 lbs. of beef when you make chili. How many lbs. Of meat will you use?

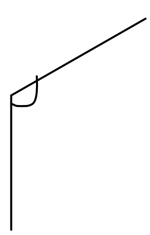
____

A4. Write the number sentence that is missing from the fact family.

9	+	5	=	14
5	+	9	=	14
14	-	5	=	9

____

G4. Choose if the angle is <, >, or = a right  angle



a) <
b) >
c) =

M4. Write <, >, or =.

10 in. ☐ 2 ft.

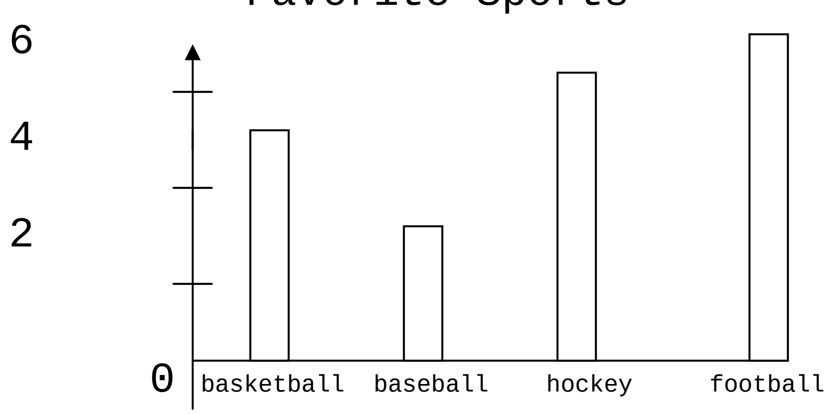
P4. Write the product that the symbol stands for.

	1	**2**	**3**	**4**
1	*1*			
2		*4*		*8*
3			*9*	
4				*

* = ________

D4. Use the bar graph.

Favorite Sports



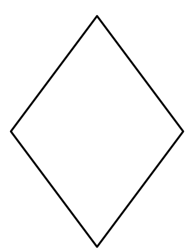
How many more students like football than baseball?________

N5. Janet wants to buy 4 notebooks for school. Each notebook cost $1.50. How much will she spend? ________

A5. Choose <, >, or =.

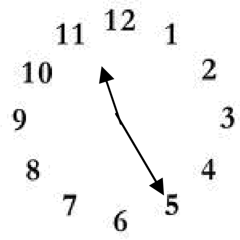
398 □ 276

G5. Find the total number of lines of symmetry for the figure.



a) 2
b) 3
c) 4

M5. Write the time.



____

P5. Write the product that the symbol stands for.

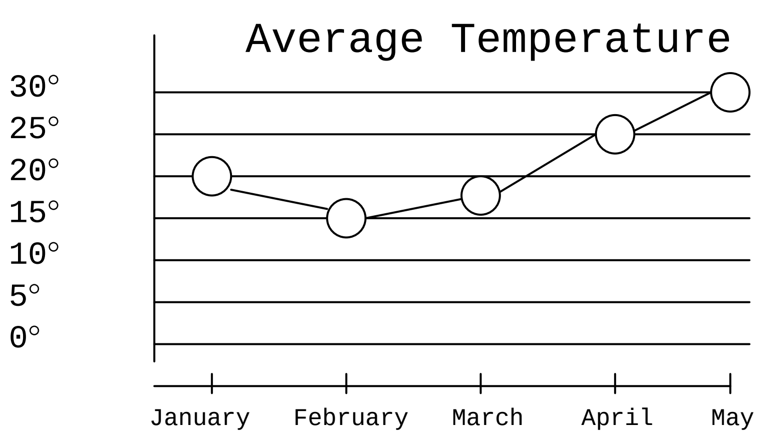
	1	2	3	4	5
1	*1*				
2		*4*			*10*
3		*6*			*15*
4		*8*		*	*20*
5			*15*		

* = ________

D5. Use the line graph.

How much colder is it in February than it is in May?

____



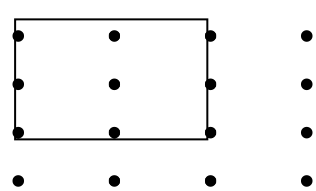
N6. Solve the problem. Decide the remainder.

A football costs $2.50. You have $8.00. How much money will you have left after you buy as many footballs as you can? ________

P6. Choose <, >, or =.

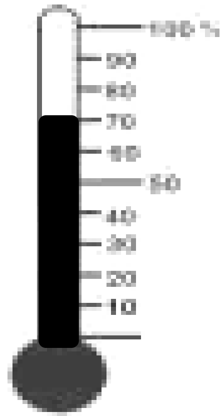
2517 ☐ 3208

D6. Choose how many.



a) 4 square units
b) 5 square units
c) 6 square units

M6. Write the temperature.



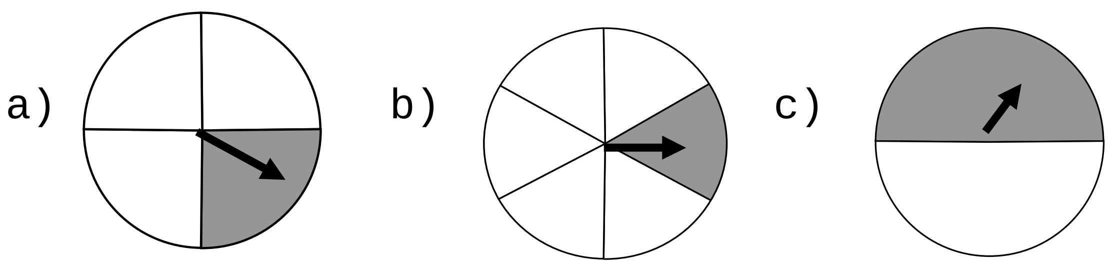
____

P6. Write the product that the symbol stands for.

	1	2	3	4	5
1		*2*			*5*
2					
3			*9*		*
4					
5		*10*			*25*

* = ________

D6. Use the spinners.



If you could win a game by stopping on a shaded area most often, which spinner would you want to use? ________

APPENDIX I

ASSESSMENT RATING PROFILE-15 (COMPUTATIONS)

Name: ____________________________________ Date: ____________________

Assessment Rating Profile (ARP – 15)

Directions: Please rate the **COMPUTATION** math assessment along the following dimensions. Please circle the number which best describes your agreement or disagreement with each statement.

	Strongly Disagree	***Disagree***	***Disagree Slightly***	***Slightly Agree***	***Agree***	***Strongly Agree***
1. This would be an acceptable assessment for assessing overall student performance in math.	1	2	3	4	5	6
2. Most teachers would find this assessment appropriate for assessing overall math performance.	1	2	3	4	5	6
3. This assessment should prove effective in monitoring student performance across the year.	1	2	3	4	5	6
4. I would suggest this assessment to other teachers who want to assess math performance	1	2	3	4	5	6
5. The student's math skills can be monitored using this assessment.	1	2	3	4	5	6
6. Most teachers would find this assessment suitable for assessing math performance.	1	2	3	4	5	6
7. I would be willing to use this assessment in a classroom setting.	1	2	3	4	5	6
8. This assessment would *not* result in negative side effects for the child.	1	2	3	4	5	6
9. This assessment would be appropriate for a variety of children.	1	2	3	4	5	6
10. This assessment is consistent with those I have used in classroom settings.	1	2	3	4	5	6
11. The assessment was a fair way to assess the content standards.	1	2	3	4	5	6
12. This assessment is reasonable for the content standards described.	1	2	3	4	5	6
13. I liked the items used in this assessment.	1	2	3	4	5	6
14. This assessment is a good way to assess math performance.	1	2	3	4	5	6
15. Overall, the assessment would be beneficial for a child.	1	2	3	4	5	6

APPENDIX J

ASSESSMENT RATING PROFILE-15 (APPLICATIONS)

Name: ____________________________________ Date: ____________________

Assessment Rating Profile (ARP – 15)

Directions: Please rate the **APPLICATIONS** math assessment along the following dimensions. Please circle the number which best describes your agreement or disagreement with each statement.

	Strongly Disagree	*Disagree*	*Disagree Slightly*	*Slightly Agree*	*Agree*	*Strongly Agree*
1. This would be an acceptable assessment for assessing overall student performance in math.	1	2	3	4	5	6
2. Most teachers would find this assessment appropriate for assessing overall math performance.	1	2	3	4	5	6
3. This assessment should prove effective in monitoring student performance across the year.	1	2	3	4	5	6
4. I would suggest this assessment to other teachers who want to assess math performance	1	2	3	4	5	6
5. The student's math skills can be monitored using this assessment.	1	2	3	4	5	6
6. Most teachers would find this assessment suitable for assessing math performance.	1	2	3	4	5	6
7. I would be willing to use this assessment in a classroom setting.	1	2	3	4	5	6
8. This assessment would *not* result in negative side effects for the child.	1	2	3	4	5	6
9. This assessment would be appropriate for a variety of children.	1	2	3	4	5	6
10. This assessment is consistent with those I have used in classroom settings.	1	2	3	4	5	6
11. The assessment was a fair way to assess the content standards.	1	2	3	4	5	6
12. This assessment is reasonable for the content standards described.	1	2	3	4	5	6
13. I liked the items used in this assessment.	1	2	3	4	5	6
14. This assessment is a good way to assess math performance.	1	2	3	4	5	6
15. Overall, the assessment would be beneficial for a child.	1	2	3	4	5	6

COPYRIGHT RESTRICTION PAGE

Due to copyright restrictions, the following manual could not be reproduced here.

Monitoring Basic Skills Progress (MBSP)

© 1998

All Rights Reserved

For more information contact the copyright holder:

Pro-Ed

8700 Shoal Creek Boulevard

Austin, TX 78757

(800) 897-3202

(800) 397-7633