Leigh M. McCulloch, A SYSTEMATIC REIVEW AND META-ANALYSIS OF CRITERION-RELATED VALIDITY IN EARLY MATHEMATICS CURRICULUM-BASED MEASUREMENT (Under the direction of Dr. Scott Methe) Department of Psychology, March 2010.

Research interest in early mathematics curriculum-based measurement (EM-CBM) has increased substantially throughout the course of the past decade. There has been a significant increase in the number of published studies regarding the validation of EM-CBM. Currently, however, there is no quantification or summarization of the multitude of research studies.

Curriculum-based measurement can be used in various ways in a school: (a) screening students, (b) monitoring progress, (c) identifying student strengths and weaknesses, and (d) predicting student performance on standardized assessments. Mathematics criterion measures are standardized, norm-referenced, individually administered tests. The purpose of administering mathematics criterion measures in the studies that will be synthesized in the proposed meta-analysis was to establish validity of early mathematics curriculum-based measurement. Mathematics criterion measures are also administered in order to measure an individual's math achievement level, as compared with same aged peers in a norm-referenced group. Finally, math criterion tests are necessary in order to define the construct that early mathematics CBM is purportedly measuring.

A meta-analysis technique was used to quantify the predictor-criterion relationship between EM-CBM and standardized norm-referenced math achievement tests. Research databases were searched to collect all relevant publications. The articles included reported correlation coefficients between EM-CBM and norm referenced standardized achievement tests, used a clear, standardized administration and scoring criteria, administered standardized math criterion assessments concurrently with, or after, the administration of the EM-CBM, and included a sample of participants in the grades between Pre-K and 2. Correlation coefficients

were obtained for each predictor-criterion relationship of interest and used as the primary units of analyses.

The first hypothesis was that there would be a strong, positive correlation between the predictor and criterion measures. The results support this hypothesis and indicate that the mean correlation between early numeracy and math achievement is .49. This correlation coefficient signifies a moderate-to-strong relationship between the two variables.

The second objective of this study was to examine the variables which influence the relationship between early numeracy and math achievement and determine which variables are moderators. There were six variables that were identified as moderators: correlation type, predictor skill, criterion skill, grade level, procedural integrity, and predictor category. Specifically, these six variables were qualitative variables found to influence the strength of the relationship between the predictor and criterion variables.

A SYSTEMATIC REIVEW AND META-ANALYSIS OF CRITERION-RELATED VALIDITY IN EARLY MATHEMATICS CURRICULUM-BASED MEASUREMENT

A Thesis Presented to

the Faculty of the Department of Psychology

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in School Psychology

by Leigh M. McCulloch

March

2010

A SYSTEMATIC REIVEW AND META-ANALYSIS OF CRITERION-RELATED VALIDITY IN EARLY MATHEMATICS CURRICULUM-BASED MEASUREMENT

by

Leigh M. McCulloch

APPROVED BY:

DIRECTOR OF THESIS

____
Scott Methe, Ph.D.

COMMITTEE MEMBER

____
Christy Walcott, Ph.D.

COMMITTEE MEMBER

____
Mark Bowler, Ph.D.

CHAIR, DEPARTMENT OF PSYCHOLOGY

____
Kathleen Row, Ph.D.

DEAN OF THE GRADUATE SCHOOL

____
Paul Gemperline, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Research interest in early mathematics curriculum-based measurement (EM-CBM) has increased substantially throughout the course of the past decade. In addition to interest, substantially more research has been conducted on this topic as well. Currently, however, no quantification or summarization of the multitude of research studies exists. One purpose of this chapter is to review the literature which addresses the development and utilization of EM-CBM. The second purpose is to discuss key early mathematical skills. An additional goal is to define the criterion measures which are used to measure achievement in mathematics. The final purpose of this chapter is to propose a method of quantifying and summarizing the current published literature on EM-CBM and math criterion measures, while examining those variables that affect their relationship.

Early Math CBM

Curriculum-based Measurement (CBM) is a “standardized methodology that specifies procedures for selecting test stimuli from a student’s curriculum, administering and scoring tests, summarizing the assessment information, and using the information to formulate instructional decisions in the basic skills areas” (Fuchs & Fuchs, 1988, p.3). These basic academic skill areas include reading, spelling, mathematics calculation, and written expression. CBM was developed in a way so that it can be administered in a cost effective and time efficient manner, thus allowing frequent use to assess indicators of basic skills (Shinn & Bamonto, 1998). Due to its sensitive nature, CBM is able to take a measurement of student performance in a basic academic skill area and quantify it in order to measure student growth over time (Fuchs, Fuchs, Bishop, & Hamlett, 1992; Hartman & Fuller, 1997).

Curriculum-based measurement can be used in various ways in a school: (a) screening students, (b) monitoring progress, (c) identifying student strengths and weaknesses, and (d) predicting student performance on standardized assessments. Using CBM for screening is useful in establishing school wide and/or local norms. It is important to establish local norms to understand normative expectations of students across the grade levels within a school district. Screening an entire school is also useful in identifying students at risk for academic problems (Naglieri & Crockett, 2005). Screening also supports intervention development for at-risk students so that they do not fall further behind their peers. Furthermore, intervening with at risk students allows practitioners to identify students who are not responding to interventions, and may prevent students who are responding positively to interventions from being miscategorized as learning disabled. Finally, CBM is used to predict student performance on standardized assessments (Crawford, Tindal, & Steiber, 2001).

Progress monitoring and tracking the progress of an individual student is another use of CBM. Progress can be tracked weekly or monthly and monitored towards long term objectives. Progress monitoring allows intervention agents to evaluate the degree of student improvement. Another important use of CBM is that it allows an instructional diagnosis, used to identify a student's strengths and weaknesses in specific academic areas (Woolfolk, 2009). Diagnosis allows teachers to set goals for their students and to match instruction to the student needs (Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 1986; Mirkin, Deno, Stecker & Fuchs, 2000; Tindal, & Kuehnle, 1982).

A general outcome measure (GOM) is a method of evaluating student progress towards long-term educational goals using reliable and valid measures tied to the curriculum (Floyd, Hojnoski & Key, 2006; Fuchs & Deno, 1991; Hintze, Christ, & Methe, 2006). Curriculum-based measurement is one of the most well-researched techniques which uses the GOM approach (Deno, 1985; Floyd et al., 2006). CBM is used to assess whether a student is making progress toward the indicated end of year goals. This is useful for teachers in both setting the objective, as well as monitoring students' growth. A varying approach is known as specific subskill mastery measurement (SSM), which is useful for diagnosis and monitoring progress toward short-term objectives (Hintze et al., 2006). Overall, GOM and SSM differ with regard to how objectives are sampled from the curriculum.

Mathematics curriculum-based measurement is a set of validated measures similar to other CBM measures in that it is completed quickly and easy to administer as well as score (Clarke & Shinn, 2004) but focuses specifically on mathematical computations. The probes are timed and primarily contain computational problems (Helwig, Anderson, & Tindal, 2002). EM-CBM is designed to be used with preschool and early elementary school students. EM-CBM focuses on basic understanding of early numeracy and contains conceptual problems rather than computational tasks.

The literature includes many studies that have examined the use of EM-CBM. Reid, Morgan, DiPerna, & Lei (2006) conducted a study to develop those measures which assess early academic skills. The authors took part in the *EARLI* project, which sought to develop screening tools that were brief and easy to use; and would be able to measure student growth in specific early academic skill competencies (Reid et al., 2006). The probes examined measurement, number identification, counting aloud, and pattern recognition, identifying the number of objects in a set, and counting objects. Results suggested that each of the math measures had adequate item and scale reliability, as well as moderate concurrent validity with selected subtests from the WJ-III test of achievement.

A similar study was conducted by VanDerHeyden, Broussard, and Cooley (2006) to further develop early math measures for preschool aged children. These authors sought to determine the accuracy and sensitivity of the constructs, and the degree to which they were able to predict later math achievement. Their measures included counting objects, selecting numbers, naming numbers, counting, and visual discrimination. Their findings suggest that performance on these measures was moderately-to-strongly correlated with kindergarten achievement. Results about the sensitivity of the early math probes to student growth were inconclusive, and therefore their use in progress monitoring is still undetermined.

Another study examined the reliability and validity of the Preschool Numeracy Indicators (PNIs; Floyd et al., 2006). The PNIs are tasks designed for preschoolers to measure their mathematics and number skills. The authors tested children aged 3 to 6 individually in classrooms. This study measured the following early math skills: One-to-one correspondence counting fluency, oral counting fluency, number naming fluency, and quantity comparison fluency. Furthermore, the PNIs were found to predict outcome on the children's performance on math criterion measures, which provides predictive validity for the measures.

Methe, Hintze, and Floyd (2008) developed brief probes designed to measure early mathematical numeracy. The authors aligned their development of assessment measures with that of the "big ideas of early mathematics" (Clements, 2004). These key ideas include: counting, comparing and ordering, equal partitioning, decomposing, grouping and place value, composing and decomposing, and addition and subtraction (Clements, 2004). These authors developed measures referred to as the Early Numeracy Skill Indicators (ENSIs). In these measures, there were four specific probes that were created: counting-on fluency (COF), ordinal position fluency (OPF), number recognition fluency (NRF), and match quantity fluency (MQF). Children enrolled in kindergarten in the northeastern United States completed these probes and a math criterion test three times throughout the year. They were also rated by their teachers on their math performance. Results suggest that the number recognition fluency and math quantity fluency measures had high reliability and validity. The number recognition fluency measure had the strongest predictive correlation with the math criterion measure. The remaining three measures had moderate predictive validity.

An additional study aimed to examine the reliability, validity, and sensitivity of four measures of early math literacy (Clarke & Shinn, 2004). In this study, the authors developed four one-minute probes designed for first grade students: oral counting measure (OC), number identification measure (NI), quantity discrimination measure (QD), and missing number measure (MN). The quantity discrimination and missing number measures had the strongest support as indicators of early mathematics skills.

Evidence from these studies on the measure of early numeracy suggests that there were similarities in the mathematical concepts that were being assessed. A number of the studies found support for the evidence of reliability and validity for the following early math measures: number identification, quantity discrimination, and identification of a missing number in a counting sequence (Foegen, Jiban, & Deno, 2007). Synthesizing research around key themes in early math assessment is a key purpose of this meta-analytic review in order to build empirical and practical confidence in the ability of the early math curriculum-based measures. Upon establishment of the importance in developing early numeracy skills, assessing these skills will be important in measuring current as well as predicted later math achievement.

Math Criterion Measures

Mathematics criterion measures are standardized, norm-referenced, individually administered tests. The purpose of administering mathematics criterion measures in the studies that are synthesized in the meta-analysis was to establish validity of early mathematics curriculum-based measurement. Mathematics criterion measures are also administered in order to measure an individual's math achievement level, as compared with same aged peers in a norm-referenced group. Finally, math criterion tests are necessary in order to define the construct that early mathematics CBM is purportedly measuring.

Some examples of math criterion tests are: (a) WJ-III Applied Problems subtest, (b) Number Knowledge Test, and (c) Test of Early Mathematics Ability, Third Edition. The Woodcock Johnson, Test of Achievement, Third Edition, Applied Problems subtest (WJ-III NU AP; Woodcock, McGrew, & Mather, 2001) is an individually administered norm-referenced test of applied mathematics consisting of problems that require the use of mathematical operations to solve a variety of applied math problems. This test is designed to measure math reasoning skills. Children are expected to answer questions regarding quantity, counting of objects, and comparison of quantities (Woodcock et al., 2001). The Number Knowledge Test (NKT; Okamoto & Case, 1996) contains four levels and students are required to obtain a minimum number of correct responses at one level to move on to the next level. The Test of Early

Mathematics Ability, Third Edition (TEMA-3; Ginsburg & Baroody, 2003) contains four categories of items that assess informal math: (a) numbering skills, (b) number-comparison facility, and (c) calculation skills. The previous tests are all used in order to assess a child's current mathematics achievement level.

Definition of Meta-Analysis

A meta-analysis is a statistical technique in which a researcher reviews the entire body of literature on a particular subject and then provides a quantitative summary of all of the findings (Cooper & Hedges, 1994; Hedges & Olkin, 1985). It has been used in studies like the current one in order to establish the validity of Reading CBM measures in its ability to predict levels of reading achievement (Swanson, Trainin, Necoechea, & Hammill, 2003). This study followed similar methodology in order to provide a quantitative summary of the findings relevant to EM-CBM and criterion measures.

Statement of the Problem and Significance of the Research

Throughout the past decade, a substantial increase in research and publications involving EM-CBM has occurred. Additionally, numerous variables are likely to moderate the relationship between EM-CBM and criterion. A moderator is a quantitative or qualitative variable that influences the strength and/or direction of the relationship between predictor and criterion variables (Baron & Kenny, 1986). Some of these hypothesized moderators in the proposed study are: age, grade level, SES, location, ethnicity, and gender. Currently, a meta-analysis of this topic has yet to be conducted, and the field is yet to be informed about a variety of issues, the most basic of which is the number of correlations available in the research. Without a summation of the abundant information in EM-CBM, some faulty conclusions regarding the relevance of EM-CBM to a mathematical construct could be made. This study will add to the literature by quantitatively summarizing the vast amount of descriptive and criterion-related validity evidence and making data-based conclusions about key technical features of common mathematics assessments. The overarching purpose of this study was to examine the nature and strength of the relationship between Early Mathematics Curriculum-based Measurement and the criterion measures of math literacy. This aided in establishing the degree to which consensus of significant extensive data exists on the utility and (predictive) validity of EM-CBM measures and early math numeracy tests. Related to this, the secondary purpose of this study was to examine the variables that may moderate the relationship between EM-CBM and criterion measures.

Hypotheses and Research Questions

The main research objective is to investigate the strength and direction of the correlational relationship between measures of early math numeracy and standardized, norm-referenced measures of math achievement. The first predicted hypothesis is that a strong, positive correlation exists between measures of early math numeracy and standardized, norm-referenced measures of math achievement. Furthermore, this study aims to determine the variables which may moderate this relationship. The second hypothesis of the study is that the overall correlation will be affected by proposed moderators, which may include one or a combination of numerous variables such as grade level, gender, location, ethnicity, SES, education status, and administration format.

CHAPTER II: REVIEW OF THE LITERATURE

Curriculum-based Assessment and Measurement

Curriculum-Based Assessment (CBA) is a set of multiple methods for sampling information about student progress relevant to curriculum goals (Hintze et al., 2006). Curriculum-Based Measurement (CBM) is a type of CBA developed in the 1970s by Deno and colleagues (Deno, 1985, 1986; Fuchs et al., 1984; Shinn, 1989), refers to a set of standardized measurement procedures designed to evaluate academic performance in reading, writing, and mathematics. As stated in Chapter 1, the focus of CBM is to monitor a student's progress towards long-term goals, whereas CBA approaches measure progress towards shorter-term objectives (Hintze et al., 2006). The sensitive nature of CBM enables the measurement of small changes in academic performance. Additionally, CBM is used to compare individual students to local or national norms (Hintze et al., 2006).

CBM was originally developed to assess student performance in reading and most research has focused on that area of assessment. However, measures have been developed to focus on additional academic areas. Recently, noted shift has been documented towards a focus on the mathematics domain. Mathematics Curriculum-Based Measurement (M-CBM) is procedurally similar to reading CBM measures; that is, it has a standardized administration and scoring, it is easy to score, and it is a "short duration fluency measure" (Clarke & Shinn, 2004, p. 235). M-CBM was designed to assess a student's performance in grades one through six. Early Mathematics Curriculum-Based Measurement (EM-CBM) is designed to assess children in Pre-Kindergarten through grade two in early mathematic skills, knowledge, and understanding. EM-CBM is necessary in addition to M-CBM because it can assess children at a younger developmental age and identify those children who may be at risk of a mathematics disability

(Berch, 2005; Clarke & Shinn, 2004). Measuring early mathematics allows for the implementation of early intervention strategies for those children who may fall behind the norm.

EM-CBM measures are currently being developed by researchers and tested for their technical adequacy. Specifically, this meta-analysis examines the concurrent and predictive validity of the EM-CBM measures. Concurrent validity displays the degree to which the EM-CBM can measure a student's performance and obtain the same results as the student's performance as assessed by another reliable measure. Predictive validity is used to demonstrate that the EM-CBM measures can predict the student's future performance on another reliable measure.

Administration and Scoring Procedures

EM-CBM was intended to be administered in an individual setting. For each EM-CBM probe, there is a standard set of directions that the examiner is instructed to read aloud to the child. Some measures require the child to write his or her answers, while other measures require the child to orally present his or her answers. Because EM-CBM is a measure of fluency, each probe is timed. Some measures are continuously timed, meaning they allow the child to complete as many problems as they can in one minute. Other measures are latency timed, meaning they record the time it takes the child to answer each question.

Typically, the main score calculated in EM-CBM probes is units correct per minute. This score reflects a child's accuracy and fluency in a skill. Acquisition, fluency, generalization, and adaptation represent the four stages that make up the learning hierarchy (Daly, Lentz, & Boyer, 1996; Eaton, & Hansen, 1978; Haring, Lovitt,). CBM measures were designed to assess a student's fluency level. Fluency refers to the student's speed of responding (Daly, Lentz, &

Boyer, 1996; Haring, et al., 1978). This score allows researchers to compare a student's accuracy and speed of responding. The goal is for children to produce accurate answers quickly.

Scoring of EM-CBM probes is calculated in one of two ways, both of which represent a fluency measure. The first way is computed by recording the number of digits the child answered correctly in one minute. The second scoring option is to record the number of problems the child answered correctly in the allotted time period. For example, if the child was given the addition problem 8 + 4 and he answered 12, the child would get 2 digits correct when using the first scoring option, and 1 problem correct using the second scoring option. When researchers use latency timing measure, scores are calculated in a similar fashion, and then converted into a digits correct per minute metric.

Importance of Measuring Early Mathematics

It is important to measure early numeracy in order to identify those students who are at risk for failure, so as to trigger an early intervention plan which will help in the long term success of these students. Prevention of children developing severe problems is the main focus of implementing interventions at an early age (Clarke & Shinn, 2004). In addition, the longer a student struggles and lags behind in mathematics, the less likely mathematics engagement and motivation are to continue (Clarke & Shinn, 2004). The National Council of Teachers of Mathematics (NCTM, 2000) conducted a literature summary and found that the "foundation for children's mathematical development is established in the earliest years" (p. 73). This suggests that intervening during the developmental years decreases the likelihood that children develop severe problems. Furthermore, it is important to develop early numeracy because it is a valid predictor of later math achievement (Jimerson, Egeland, & Teo, 1999; Stevenson & Newman, 1986). Because of the importance of early intervention, the measures being used to assess children must reliably achieve three objectives: (1) identification of at-risk students, (2) monitoring of students' growth in those skills in which they are at risk for failing, and (3) consistence of the accuracy of scores across raters, form, and time (Clarke & Shinn, 2004).

Early Numeracy

According to many authors, early numeracy is a broad term describing number sense (Chard et al., 2005; Clarke & Shinn, 2004; Clements, 2004; Clements et al., 2004; Berch, 2005). Early mathematics describes the development of children as they enter their school years and acquire both nonnumeric and numeric knowledge (Baroody, 2004). Number sense is defined differently by researchers, but in the literature it generally refers to certain mathematical abilities among children. Number sense is a pre-requisite to learning mathematical skills, and therefore tends to be acquired in the preschool years. Berch (2005) examined the research of over a dozen articles relating to number sense and defines it as:

> Possessing number sense ostensibly permits one to achieve everything from understanding the meaning of numbers to developing strategies for solving complex math problems; from making simple magnitude comparisons to inventing procedures for conducing numerical operations; and from recognizing gross numerical errors to using quantitative methods for communicating processing, and interpreting information. (p. 333-334).

Debate still exists amongst experts in the field as to whether number sense is an innate ability or whether it is a skill that is taught and developed through repeated practice. However, agreement exists that children first acquire informal and conceptual knowledge, that is, knowledge about enumeration, counting, size, position, and decomposition (Clements et al., 2004; Russell & Ginsburg, 1984). Following informal knowledge, children begin to develop formal and numeral based knowledge (Russell & Ginsburg, 1984). This type of knowledge involves the use of the actual written representations of numbers.

Early Numeracy Concepts

Seven skill categories are used to characterize the early numeracy assessment tools named in the articles analyzed in the current study. These skill categories were determined based on the comprehensive work and research of Clements, Sarama, and DiBiase (2004) in combination with the coding process used in the current study. The seven skill categories are as follows: (a) readiness concepts, (b) counting, (c) comparing and ordering, (d) equal partitioning, (e) composing and decomposing, (f) grouping and place value, and (g) adding to/taking away.

Readiness concepts. Readiness concepts refer to informal measures of numeracy including knowledge about shapes and colors, language concepts pertaining to size, and pattern recognition. Readiness skills as measured by the Bracken Basic Concept Scale - Third Edition (BBCS-3) target receptive language skills and basic concept acquisition and knowledge (Bracken, 2006). The math readiness concepts that are measured by the BBCS-R include colors, direction/position, textiles/materials, time, sizes, and shapes (Bracken, 2006). Math readiness concepts are used as an indicator of the child's ability to acquire formal math skills and concepts. Counting can be measured using both informal and formal measures. Children competent in counting must be able to use the number names to count and recognize the connection between the quantity and the word (Baroody, 2004). Children begin by stringing number words together (Baroody, 2004) then count by ones in numerical order. Next, children learn to begin counting at various points other than the number one. A final step is the ability to count backwards (Baroody, 2004).

Comparing and ordering. Comparing and ordering involves comparing groups of items, determining which one is bigger, smaller, or whether the quantities are equal. This category also includes the skill of identifying the order or position of people, places, or things, and naming their rank order, such as first, second, and third in a race. An example of comparing is showing a child two pictures of colored circles and asking the child to select the picture that has more circles.

Equal partitioning. Equal partitioning refers to the ability to separate a group of things into sets of equal sizes. Yoshida and Sawano (2002) refer to equal partitioning as an informal knowledge of breaking apart a whole into equal parts, or half of a whole. Children have typically acquired early knowledge of this skill before entering kindergarten (Nunes & Bryant, 1996). For example, a child given ten cookies can give the same number of cookies to himself and a friend.

Composing and decomposing. Composing and decomposing is the awareness that a whole can be composed from or decomposed into different parts (Baroody, 2004, p. 209). This category is analogous to equal partitioning, but this category requires children to deal with "larger numbers in a more abstract fashion" (Methe & Riley-Tillman, 2008, p. 32). Equal partitioning requires that children separate a whole into equal parts, while decomposing entails identifying several ways of breaking down the same whole. A child recognizes that the ten cookies can be made up of seven and three, five and five, two and eight, one and nine, and four and six.

Grouping and place value. Grouping and place value involves grouping items into larger units. For example, if asked to count 100 pennies, a child would group the pennies into separate categories in order to assist them with their counting. Place value refers to the understanding that the value of each digit in a multi-digit number corresponds to the position of the digit (Baroody, 2004). This requires knowledge of the ones place, tens place, hundreds place, etc.

Adding to and taking away. Adding to and taking away refers to the idea that children are able to recognize the addition to and subtraction from groups of items (Methe & Riley-Tillman, 2008). This also involves the ability to complete written addition and subtraction problems. It is suggested that this ability incorporates the earlier skills; but that children do not rely on these strategies and develop more efficient ones (Methe & Riley-Tillman, 2008). For example, when a child is asked to add twelve and six, instead of using his fingers to count, the child must first add the two and six, then add one plus zero.

Predictor Measures

Methe, Hintze, and Floyd (2008) constructed four measures of early numeracy and referred to them as the Early Numeracy Skill Indicators (ENSIs). The four measures are the following: Counting-on Fluency (COF), Ordinal Position Fluency (OPF), Number Recognition Fluency (NRF), and Match Quantity Fluency (MQF). COF was developed to measure a child's counting ability. OPF was designed to assess a child's ability to determine the order, or place of objects. NRF required children to accurately name written numerals as quickly as possible. Finally, MQF required children to match a pictorial description of a group of objects with the corresponding written numeral. The reliability of these measures was reported as follows: COF (.80), MQF (.53), and OPF (.83).

Floyd, Hojnoski, and Key (2006) developed and experimented with four early numeracy measures called the Preschool Numeracy Indicators (PNIs). The first is called one-to-one correspondence counting fluency. This probe, designed to measure the ability to fluently count objects (Floyd, Hojnoski, & Key, 2006), requires children to count the number of objects on a page as quickly as possible. The second probe, called oral counting fluency, requires children to count aloud in sequence starting with the number one (Floyd et al., 2006). The third measure, number naming fluency, requires children to accurately name a written numeral as quickly as possible when it is presented in random order (Floyd et al.,2006).   The final measure Floyd and colleagues used was the quantity comparison fluency probe. In this measure children identify the larger of two groups of objects as quickly as possible (Floyd et al., 2006). Reliability data were not reported for these measures.

A number of researchers have used experimental early numeracy probes that are analogous to the aforementioned ones, but using their own specific name. For example, Clarke, Baker, Smolkowski, and Chard (2008) used measures named Oral Counting (OC), Number Identification (NI), Quantity Discrimination (QD), and Missing Number (MN). It is evident that these measures correspond to the previous measures, and that they assess the same early numeracy skill. Fuchs, Hamlett, and Fuchs developed progress monitoring measures for elementary mathematics called the Monitoring Basic Skills Progress measures (MBSP) (Foegen et al., 2007). Two types of MBSP measures exist. The first is referred to as computation probes. Thirty comparable forms have been designed for grades one through six. These measures were developed by Fuchs, Hamlett, and Fuchs selecting computation problems that were integrated into the curriculum at each grade level in Tennessee (Foegen et al., 2007). The second type of MBSP measure is referred to as concepts and application probes, which were designed for children in grades two through six. These probes were developed to assess mathematical concepts and applied skills such as solving word problems and understanding charts and graphs (Foegen et al., 2007).

Development of Mathematics Skills

Several theories regarding the progression of the development of mathematics skills have been established. However, there is agreement that mathematics is a broad term that encompasses a variety of skills, concepts, and knowledge. According to Aunola, Leskinen, Lerkkanen, and Nurmi (2004), math ability ranges between a basic understanding of the number system, rote memorization of facts, and creating solutions to increasingly complex problems. Furthermore, it is hypothesized that mathematical skills are developed in a hierarchical order, suggesting that understanding and solving abstract problems requires that a student first become fluent in the more basic skills (Baroody, 2005; Aunola et al., 2004)

Baroody (2005) presents a theory of early mathematics skill development called the Mental Models View, which suggests that three key stages exist that young children transition through in their development of mathematical competence. The three phases are as follows: (1) Transition 1: The development of exact pre-counting numerical and arithmetic process, (2) Transition 2: The development of counting-based numerical and arithmetic competencies, and (3) Transition 3: the development of written representations (Baroody, 2005, p. 175; Huttenlocher, Jordan, & Levin, 1994; Mix, Huttenlocher, & Levine, 2002). It is suggested that for a child to fully understand the basic meaning of numbers, he or she must understand that numbers have four distinct meanings. Baroody (2005) described these meanings of numbers using the following questions: (a) how many? (b) how much? (c) where? (d) what? The question "how many?" refers to the cardinal meaning of the number (Baroody, 2005), that is the number four has a direct meaning of four objects. The question "how much?" is used to represent size. How much bigger is one ball than another ball? The question "where?" refers to position, rank, or order, such as the rank of people finishing a race. The final question used to represent the role of a number, “what?”refers to non-mathematical uses of numbers. For example, in sports, each teammate is assigned a different number. These numbers have no meaning other than identification.

In addition to these numerical and arithmetic competencies of early numeracy proposed by Huttenlocher, Jordan, and Levin (1994), the National Research Council (NRC) categorized the progression of mathematical achievement into five groups. They note that the groups are interdependent of each other (NRC, 2001). The categories are the following: (a) conceptual understanding, (b) procedural fluency, (c) strategic competence, (d), adaptive reasoning, and (e) productive disposition. The NRC (2001) refers to these categories as strands, and defines them in the following way:

> Conceptual understanding refers to the comprehension of mathematical concepts, operations, and relations. Procedural fluency refers to skill in carrying out procedures flexibly, accurately, efficiently, and appropriately. Strategic competence refers to the ability to formulate, represent, and solve mathematical problems. Adaptive reasoning refers to the capacity for logical thought, reflection, explanation, and justification. Finally, productive disposition is defined as habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one’s own efficacy. (p. 116).

Both theories suggest that early numeracy is important, and must be fully developed before children are able to learn more complex mathematical procedures.

Importance of Mathematics Competence

The National Assessment of Educational Progress (NAEP) is a nationally representative assessment administered every two years in order to measure the knowledge of students across the United States in various subject areas. The most recent mathematics assessment took place between January and March, 2007, and was administered in almost 15,000 schools throughout America. Approximately 197,700 fourth grade students and 153,000 eighth grade students took the NAEP mathematics assessment (National Center for Education Statistics, 2009).

The 2007 NAEP results in mathematics revealed that fourth-graders' mathematical skills have demonstrated an increasing trend over the course of the past seventeen years. Fourth grade students scored two points higher in 2007 than they did in 2005 and 27 points higher than in 1990 (National Center for Education Statistics, 2009). Although math scores are improving, it is still an area for concern. In 2007, 18% of fourth graders scored below the basic level of achievement. Of those 82% who scored at or above the basic level of achievement, only 39% scored at or above the proficient level (National Center for Education Statistics, 2009).

In 2007, 29% of eighth grade students scored below the basic level of achievement. Of those 71% who scored at or above the basic level of achievement, only 32% scored at or above the proficient level of achievement (National Center for Education Statistics, 2009). The basic level of proficiency refers to "partial mastery" of skills and knowledge that are necessary for each grade level. The proficient level refers to "solid academic performance" and "competency over challenging subject matter" in each grade level (National Center for Education Statistics, 2009, p. 6). These results suggest that only a small percentage of students have achieved math proficiency.

According to Geary (1996), low math achievement among children in the United States is a cause for concern among many adults in the United States. Children most at risk for poor performance in mathematics are children of low SES and females (Eccles, 1997). This is disconcerting because jobs related to competence in mathematics tend to be male dominated and higher paid. Individuals who evidence mathematics proficiency in their jobs earn 38% more than other individuals (Riley, 1997). Moreover, advancing technology often demands greater math proficiency.

Mazzocco and Thompson (2005) indicate that, "poor math achievement is a risk factor for negative outcomes in both childhood and adulthood" (p. 142). This stresses the importance of math education in childhood. Math performance level has also been directly linked to the educational level an adult will achieve (Delazer, Girelli, Grana, & Domahs, 2003). Math achievement disparities beginning in childhood can result in poorer chances for full employment and well-paying jobs (Rivera-Batiz, 1992). Along with several other factors, math literacy increases one's chances of full-time employment (Rivera-Batiz, 1992). Because math is becoming such an important factor in one's future career opportunities, it is important to develop mathematics achievement at an early age. As previously stated, early numeracy is the first step to developing mathematics achievement. Hopefully, successful early intervention can help to reduce the gender and SES gap in many science and technology related careers.

Chapter Summary: The Problem and Proposed Solutions

To summarize, a number of short duration, single skill probes are used to measure early mathematical skills. Several different measures have been developed which assess the same skill categories. However, no consensus exists regarding the technical adequacy and criterion related validity of each of these measures. The purpose of this study is to systematically review the current research of the development of EM-CBM. The second purpose is to aggregate the predictive and concurrent validity data of the independent EM-CBM measures and standardized, norm referenced mathematics achievement assessments. A final purpose is to identify the impact of potential moderating variables that influence the relationship between EM-CBM and math achievement tests. To date, there has not been any research measuring the influence of the moderating variables between EM-CBM and math achievement tests. In conclusion, by validating curriculum-based measures of early numeracy practitioners would be able to assess mathematical skills and abilities in a cost and time efficient manner, develop and implement interventions in early childhood, and monitor students' performance gains in the classroom.

CHAPTER III: METHOD

Inclusion and Exclusion Criteria

In anticipation of upcoming changes to the American Psychological Association (APA) research reporting standards, the following research design and methodology adheres closely to the meta-analytic structure outlined in Cooper, Appelbaum, Maxwell, Stone, and Sher (2008). Mathematics curriculum-based measures can be defined operationally as a set of short-duration (usually one-minute) tests, the design of which lends itself to the development of alternate forms, and are used to assess an individual's current level of academic proficiency and to monitor student progress. Mathematics criterion assessments chosen for the current review are typically lengthier published standardized tests in which a child's academic performance is compared to his or her same aged peers based on a normative sample.

Criterion validity studies of EM-CBM and other standardized measures of math achievement were included in this meta-analysis. Dissertations and published research articles were examined to determine if inclusion criteria were met. Four criteria were used to determine the inclusion of articles in the overall analysis. The criteria that were used are: (a) correlation coefficients between EM-CBM and norm referenced standardized achievement tests must be available; (b) a clear, standardized administration and scoring criteria must be used; (c) standardized math criterion assessments must have been conducted after, or concurrently with, the administration of EM-CBM measures; and (d) the participants must be in grades Pre-K through 2, or between the ages 3 and 7. Given that established, peer-reviewed research is a useful primary means to determine the utility of EM-CBM, conference data or research in progress were not included in the search. Studies were also excluded in which the authors used the same sample for more than one publication. This is to ensure that each sample was represented in the overall meta-analytic sample only once.

Coding Overview

Article coding is necessary so that each separate study is comparable to the others over a set of criterion discussed below. It is essential to ensure that the articles include standardized reviews of the same variables in the same measured units, in order to facilitate analyses and reach conclusions. Coding the articles aids in transformation of the units of each variable so that the variables match each other. Only after variables are recorded in the same units can a comparison between variables take place. For example, it is beneficial to compare students in special education to other students in special education rather than comparing students in special education to students in a general education classroom since that the type of instruction varies within the two separate settings.

As in other published meta-analyses (Swanson, et al., 2003) the original sets of predictor and criterion measures, as well as several hypothesized moderating variables were coded. When possible, sub-group correlations were analyzed in addition to whole group comparisons. Further analysis of the whole group correlations also contributed to the overall analysis. Components of the following general categories were examined and coded for information relating to: (a) identifying variables (b) content of the predictor measures, (c) content of the criterion measures, (d) demographics, and (e) procedural and process variables (e.g., time of administration, administration format, etc.). A main goal was to identify these variables and code for them so that we are able to determine the effect they have on the overall predictor-criterion relationship. The following list provides the variables that were coded:

Identifying Variables

Each article was assigned a unique identification number using the basic ordinal sequence from 1-18. Within each article, each correlation identified was assigned a unique ordinal number. For example, Clarke and Shinn (2004) contained 40 reported correlation coefficients that were included in the article. Each of those correlations received its own correlation number. Each correlation coefficient was recorded with two decimal places as reported in each article. Each correlation was coded based upon the correlation type; as either a concurrent or predictive measure. Concurrent correlations represented instances in which predictor and criterion measures were administered within three weeks. Predictive correlations refer to those instances in which the predictor measure was administered at least one month prior to the administration of the criterion measure.

To address issues of restricted-range correlation coefficients due to developmental level, a hypothesized process moderator is the time of year that the predictive measures were given. For example, children who were administered the number identification EM-CBM measure in the fall of kindergarten may have had higher rates of low scores, affecting the correlation coefficients. Correlation coefficients are most aptly obtained when the spread of scores approximates a normal distribution. Although corrections for range restriction should attenuate the influence of this artifact, time is seen as a key developmental variable. The time of year the predictor measure was administered was recorded as fall, winter, or spring. The months between August and November were coded as fall. Winter was recorded when the early numeracy probes were administered between the months of December and March. Between the months of April and June, the time of predictive measure was recorded as spring.

Predictor Measures

Each measure was first assigned an abbreviation, as listed in Appendix A. Second, the skill of the predictor was identified, and differentiated along the lines of Clements (2004) recommendations. Each CBM probe has been developed to capture some important component of the early mathematics construct. A variety of skills were assessed, as the probes were developed to measure each skill in a hierarchical framework. Each predictive measure was placed into the category of the specific skill it is intended to measure. The different predictor skill types were: counting, comparing and ordering, equal partitioning, composing and decomposing, grouping and place value, addition to and taking away, shapes or other readiness concepts (color, size, patterns, etc.), and mixed or other; as identified in both Appendix A and C. Some probes were developed to measure a combination of skills. In these instances, the probe was coded as mixed skill.

In regards to predictor category, EM-CBM probes were coded as informal if they measured conceptual knowledge, counting, size, position, and decomposition (Clements et al., 2004; Russell & Ginsburg, 1984). Probes were coded as formal if they measured numeral based knowledge, involving the use of written numerals (Russell & Ginsburg, 1984). The time of each early numeracy probe was recorded. Many early numeracy measures are timed, allowing a child to complete as many problems during a predetermined time period. The measures were coded as one minute, more than one minute, but less than 4 minutes, or greater than or equal to four minutes.

When reported in the article, the mean and standard deviation of each measure were recorded. The mean represents the average score on each early numeracy probe taken by the participants in the study. The standard deviation refers to the amount of variation between scores and the average or mean score. The administration format was coded in terms of whether measures were administered in an individual or group setting. Clarke & Shinn (2004) indicated that their test battery, the TEN, was administered to individual children in a one-to-one setting. Individual administration is identified by the presence of only one examiner and one child at a time, where the test is administered according to directions in the presence of the student. Group format can be identified by one examiner with a large group of children. The examiner delivers the directions of the assessment to several students at the same time.

Each predictor measure was also coded to represent the way in which the test was timed. Measures were coded as either latency timing or continuous timing. Latency timing refers to measures that time how long it takes a child to complete each independent task; whereas continuous timing refers to measures that allow a child a predetermined amount of time to complete an entire probe. Reliability data of each measure was also recorded when reported in the article. The following types of reliability were recorded: internal consistency, alternate-form reliability, test-retest reliability, and inter-rater reliability.

Criterion Measures

Each criterion measure was also assigned an alphabetic abbreviate, as listed in Appendix B. Upon completion of gathering the studies that were input into the analysis, each subtest was grouped into and coded for the skill category it purports to assess. Each criterion measure was coded into one of the following categories: basic readiness concepts, basic math concepts, applications/problem solving, calculation of non-basic facts, or mixed skill. When reported in the article, the mean and standard deviation of each measure were recorded. Criterion score type refers to whether the math achievement measure was a single sub-test or just part of an achievement battery, a composite score derived from several subtests, or a single skill test.

Criterion norm type refers to the standardization sample of the math achievement assessment. It was recorded whether the achievement measure was nationally normed or a state-specific test with state/local norms. Similar to the predictor administration format, the administration format of the criterion measure was recorded and coded as either individual administration or group administration.

The time of administration of the criterion measure was also recorded in the same way as the predictor measures. Administration of the math achievement assessment during the months of August and November was coded as Fall. Administration between December and March was recorded as Winter, and administration between April and June was recorded as Spring. Reliability data of each measure was also recorded when reported in the article. The following types of reliability were recorded: internal consistency and/or alpha coefficient for the measure, alternate-form reliability, test-retest reliability, and inter-rater reliability.

Demographic Variables

The size of each sample that contributes to the overall correlation coefficient was recorded. This was recorded in order to calculate group averages that were corrected for sampling error. The mean grade level (grades Pre-K through 2) for the sample at the time of administration of EM-CBM measures was recorded. For studies that included age in place of grade level, grade level was based on mean age (3 and 4 years = Pre-K, 5 years = kindergarten, 6 years = first grade; 7 years = second grade).

Ethnicity was recorded based upon the numbers of participants who were Caucasian, African American, Asian American, Native American, and Hispanic in the reported samples. Samples that were made up of at least 50% of the same ethnicity were recorded as the corresponding ethnicity. Samples that were not made up of 50% of the same ethnicity were recorded as mixed ethnicity. Gender was recorded based upon predominance. Samples that had greater than 50% males were coded as male. Similarly, samples that included greater than 50% females were recorded as female. Samples that were not represented by a gender majority were coded as mixed. Additionally, the numbers of male and female students in the studies were recorded.

Free and reduced-priced lunch was used as an indicator of socioeconomic status (SES). SES was also recorded based on percentage of students eligible for free or reduced- priced lunch, and those ineligible for free or reduced-priced lunch in the school system was coded. Socioeconomic status was also recorded as predominance of a single group. As similarly recorded in McBee (2006), low SES refers to students who received either free or reduced-priced lunches and high SES refers to students who are not receiving lunch aid. Oftentimes, authors indicate in an article the percentage of students who qualify for free or reduced-price lunch when relating the descriptive results of the included sample. It should be noted that some articles did not report this variable. In these instances, this variable was recorded as “cannot tell”.

When reported, the percentage of students served in general education, special education, and English as a Second Language classrooms in the sample were coded. Education status was similarly coded based on predominance of group. These data may have overlapped with SES data. The location of where the sample resides as Mideast, Midwest, Northeast, Northwest, Southeast, and Southwest was coded when possible as based upon the wording in each article. Samples that were from outside the United States were not included in the analysis.

Procedural Variables

The type of sample was recorded as reported by the author of each article. Sample types included random samples, stratified samples, and convenience samples. Random samples are those in which a subset of individuals was selected from the population. In a random sample, each individual in the population has the same probability of being chosen and the individuals chosen for the sample were selected randomly and by chance (Shaughnessy, Zechmeister, & Zechmeister, 2008). Stratified samples are those in which participants are identified by selecting independent samples from a larger subpopulation or group within the population (Shaughnessy et al., 2008). Convenience samples are selected through including those individuals who are readily available and convenient (Shaughnessy et al., 2008). An example of a convenience sample is a researcher collecting data from the first grade students in a local school district. Generalizations about an entire population cannot be made from studying a convenience sample (Shaughnessy et al., 2008).

It was recorded whether the examiners of the predictor and criterion measures were trained as part of the research study, before administration of assessments. This was a dichotomous variable, and coded as either trained or not trained. Articles that were recorded as "examiners trained" must have specifically stated the measures that were taken to train the examiners before conducting assessments with the children in the research study. Training examiners includes, but is not limited to, instructing examiners on administration standards and scoring criteria for each early numeracy measure and standardized achievement assessment.

The qualifications of the examiners in each study were recorded. Categories included pre-professional such as graduate level student, professional such as teacher, counselor, speech and language pathologist, other, mixed, and cannot tell. Procedural integrity was measured by reported whether data, narrative, or neither was presented in an article to discuss measures to ensure the accuracy of a dataset was coded.

Search Strategies

Using the search criterion detailed below, the existing research literature was searched to identify articles to be included in the meta-analysis. Databases searched include: PsycINFO, ERIC, and Education Research Complete. The key search terms used were: CBM, mathematics CBM, early math, math tests, early numeracy, numeracy, math criterion measures, mathematics assessments, and mathematics progress monitoring. There was no limitation on the time period of publication of the identified articles.

Similar to the work of Swanson, Trainin, Necoechea, and Hammill (2003), potential articles were identified by searching for articles written by authors known to have conducted research in mathematics CBM. The reference lists of the articles identified in the initial literature search was examined to identify additional articles.

Upon retrieval of the initial articles, eligibility was established through a number of processes. First, the reviewer searched the title for any of the previously mentioned key terms that were used for the search criteria. This step was useful in excluding articles that used the acronym CBM to refer to a meaning other than curriculum-based measurement. Furthermore, articles were excluded in this initial step when the title of an article referred to curriculum-based measurement that did not assess mathematics, such as reading and writing.

The reviewer then scanned through the abstract to determine the relevance of the study. This step was important in ruling out articles. The articles most often excluded through this process were due to inappropriate age of the sample, statistics other than correlations reported, and no standardized criterion measure administered.

Finally, the reviewer sorted through the full text of the article to identify whether the article met all of the inclusion criteria, while sifting out those articles that met any of the exclusionary criteria. When the full text of the article was not available, the Inter-Library Loan process was used in order to obtain a copy of the article. There were a total of two dissertations that were not accessible to the author of this study.

There was only one reviewer judging the eligibility of the studies. However, in instances when inclusion or exclusion was ambiguous, the reviewer and her university advisor reviewed the article, discussed eligibility criteria, and made a joint decision. Using this process, articles were condensed from an initial total of 2,265 hits using the key search terms on the online databases to a total of 16 articles included in the current meta-analysis.

Coding Procedures

Each article was coded by two independent coders in order to ensure coding accuracy. The first coder coded each of the articles. One of the independent coders was a graduate student at East Carolina University enrolled in the school psychology program. As supported in several previous meta-analyses (Bear, Minke, & Manning, 2002; Rohling, Beverly, Faust, & Demakis, 2009), a second coder coded 100% of the articles. The second coder was a professor at East Carolina University in the school psychology program. Once the two coders separately coded the articles, both individuals met to discuss their coding results and determine the agreement rates between the two coders. Inter-coder agreement was calculated using the following formula: total number of agreements divided by the number of total items possible, multiplied by 100. Kappa coefficients were used when applicable to address chance agreement. Initial inter-coder agreement ranged between 85% and 95%, with an average of 92%. Upon instances of disagreement between the two coders, each coder discussed the reasoning for the code and referenced the original article for proof. The coders then came to an agreement on the appropriate code together. The goal of reaching an inter-coder agreement level exceeding 90% for the entire dataset was obtained.

Once coding was completed, data from each coded article were entered into an excel document. To complete this process there were two students from East Carolina University working together in order to ensure accuracy of data entry. The main author read the codes aloud to an undergraduate student while she entered them in the computer. While that data were being entered, the main author was watching the computer screen to ensure that the accurate numbers were being entered. Following this, the second coder checked at random 80% of the data on the excel spreadsheet.

Statistical Analytic Plan

Data collected from the coding process were entered into a spreadsheet, with top row headings used to identify the coded components for each of the inclusive studies. Correlation coefficients were obtained for each predictor-criterion relationship of interest and entered into the spreadsheet for use as the primary units of analyses. Hunter and Schmidt (2004) indicate that the primary utility of meta-analyses is to correct for both sampling error and error introduced through measurement artifacts. To correct for sampling error, (a) the variance of each set of correlations was derived and weighted by sample size and (b) expected sampling error variance for each correlation was derived. The ratio of variance expected to observed (actual) variance was computed to arrive at a percentage of variance in correlations that was accounted for by sampling error. This percentage is useful to establish the degree to which correlations are constant across moderating variables. To correct for the second type of error, correlations were corrected for restrictions in range and attenuation (Hunter & Schmidt, 2004).

CHAPTER IV: RESULTS

Study Characteristics

A total of 321 correlations were analyzed as reported from sixteen independent studies on early numeracy. Table 1 summarizes the reliability and validity of the studies that were included in this meta-analysis. Four of the articles included in this meta-analysis are represented in a similar table in Foegen, Jiban, and Deno (2007) and therefore not included in Table 1 to reduce redundancy. To aid in the understanding of Table 1, the abbreviations of the EM-CBM probes are listed in Appendix A.

The majority of students included in the samples were in kindergarten (41%). Sixty-six percent of the participants included in the samples were predominantly (≥50%) Caucasian, eleven percent were predominantly African American, and twenty-three percent of the samples were not composed of participants from a predominant ethnicity. Six of the studies failed to report the gender make-up of the participants included in the sample. Of those articles that reported the number of males and females, there were 8,609 male and 7,792 female participants.

A sample needed to consist of at least 50% of the population of the study in order to be coded for. Seven of the total sixteen articles reported SES of the participants. SES was recorded as a dichotomous variable based on predominance. Of those seven, 38% of the samples were predominantely eligible for free or reduced lunch. 77% of the studies were conducted on predominantly general education students.

Out of the six possible locations, the majority of the samples were located in the Southeastern part of the United States (31%), followed by the Northeast (25%), Midwest (19%), Northwest (19%), and cannot tell (6%). Three of the studies reported random selection of participants and one study indicated the use of a convenience sample. The majority of studies failed to indicate the specific participant selection process that derived the sample.

In all but one study, the training of examiners was documented. The qualifications of the examiners varied between pre-professionals, such as graduate students (29%), professionals, such as teachers (24%), other, such as undergraduate students (24%), mixed (22%), and 1% was not reported. Twenty-five percent of authors reported taking measures to ensure procedural integrity and presented this information using data. Another twenty-five percent of authors used narrative to describe procedural integrity precautions. Six percent of authors claimed to possess procedural integrity, but did not provide any data and 44% of the articles failed to describe any type of procedural integrity of the data collection.

Overall Correlation Coefficient

The strength and direction of the correlational relationship between measures of early numeracy and math achievement were computed. This correlation was obtained by taking each of the 321 total correlations and independently multiplying the correlation coefficient by its sample size and then taking the sum of all 321 correlations. This total sum was then divided by the total number of participants across all of the inclusive studies, which resulted in the overall mean correlation coefficient for all EM-CBM measures with the criterion measures, $r = .49$. This correlation was the result of correcting for any sampling error that may have been introduced into the meta-analysis through the culmination of research studies. The correction for sampling error was necessary because sampling error is the simple and common explanation for differences between effect sizes (Cooper, 1998).

Descriptive Data

Table 2 presents the descriptive results of variables that were coded. This demonstrates the overall composition of early numeracy measures and math achievement measures that were included across the sixteen articles. Results include the number and percent for each variable out of the total number of correlations included, corrected mean of each variable, and the uncorrected mean, standard deviation, and 95% confidence interval for each variable.

Table 2. *Descriptive Results of Measured Variables*

Moderator	Number of Correlations	Percent of Total	Corrected Mean	Uncorrected Mean (SD) [Mean +/- 95% CI]
Correlation Type				
Concurrent	209	65.1	.48	.43 (.23) [.40 - .46]
Predictive	112	34.9	.51	.50 (.18) [.46 - .53]
Predictor Administration Time				
Fall	143	44.5	.48	.47 (.19) [.44 - .50]
Winter	24	7.5	.55	.53 (.12) [.48 - .58]
Spring	154	48.0	.42	.50 (.24) [.46 - .54]
Predictor Measure Probe Time				
Cannot Tell	46	14.3	.20	.30 (.27) [.22 - .38]
One Minute	249	77.6	.50	.51 (.16) [.49 - .53]
>one minute, <four minutes	26	8.1	.46	.36 (.23) [-.52 - 1.24]
Predictor Administration Format				
Individual	259	80.7	.45	.49 (.22) [.46 - .52]
Group	62	19.3	.46	.48 (.18) [.44 - .52]
Predictor Skill				
Counting	127	39.6	.46	.47 (.16) [.44 - .50]
Comparing and Ordering	41	12.8	.57	.56 (.14) [.51 - .60]
Adding to/Taking Away	34	10.6	.51	.49 (.18) [.43 - .55]
Mixed Skill	99	30.8	.55	.46 (.21) [.42 - .51]
Readiness Concepts	20	6.2	.06	-.01 (.21)[-.11- .09]

Criterion Measure Assessment				
Cannot Tell	6	1.9	.49	.34 (.22) [.17 - .51]
Single Sub-test	120	37.4	.42	.52 (.27) [.47 - .57]
Composite Score	72	22.4	.49	.52 (.16) [.48 - .56]
Single Skill Test	123	38.3	.47	.45 (.18) [.42 - .48]
Criterion Score Type				
Cannot Tell	6	1.9	.57	.43 (.27) [.22 - .64]
Transformed Score	197	61.4	.43	.49 (.23) [.46 - .52]
Raw Score	118	36.8	.48	.49 (.18) [.46 - .52]
Criterion Administration Format				
Cannot Tell	7	2.2	.26	.26 (.12) [.17 - .35]
Individual	242	75.4	.45	.49 (.22) [.46 - .52]
Group	72	22.4	.48	.54 (.20) [.49 - .59]
Criterion Administration Time				
Cannot Tell	43	13.4	.20	.20 (.28) [.12 - .28]
Fall	57	17.8	.48	.47 (.20) [.42 - .52]
Winter	31	9.7	.42	.45 (.22) [.37 - .53]
Spring	190	59.2	.51	.51 (.15) [.49 - .53]

Moderator	Number of Correlations	Percent of Total	Corrected Mean	Uncorrected Mean (SD) [Mean +/- 95% CI]
Criterion Skill				
Cannot tell	14	4.4	.47	.48 (.14) [.40 - .56]
Basic Readiness Concepts	4	1.2	.51	.51 (.10) [.35 - .66]
Basic Math Concepts	125	38.9	.45	.44 (.17) [.41 - .47]
Applications/Problem Solving	60	18.7	.41	.38 (.30) [.30 - .46]
Calculation of Non-basic Facts	31	9.7	.54	.60 (.16) [.54 - .66]
Mixed Skill	87	27.1	.53	.46 (.21) [.42 - .51]
Grade Level				
Pre-Kindergarten (PK)	10	3.1	.41	.41 (.16) [.29 - .52]
Kindergarten (K)	131	40.8	.50	.39 (.23) [.35 - .43]
First (F)	115	35.8	.44	.49 (.22) [.45 - .53]
Second (S)	8	2.5	.56	.56 (.18) [.41 - .70]
Mixed (M)	57	17.8	.54	.52 (.11) [.49 - .55]
Examiner Training				
Cannot tell (CT)	16	5.0	.48	.48 (.09) [.43 - .52]
Trained (T)	305	95.0	.49	.45 (.22) [.43 - .48]
Procedural Integrity				
Cannot tell (CT)	104	32.4	.34	.32 (.23) [.27 - .37]
Used Data (UD)	108	33.6	.54	.52 (.21) [.48 - .56]
No Data (ND)	5	1.6	.49	.49 (.11) [.35 - .63]
Narrative (N)	104	32.4	.52	.52 (.14) [.49 - .54]
Predictor Category				
Cannot tell (CT)	6	1.9	.70	.70 (.02) [.68 - .73]
Informal (I)	131	40.8	.45	.39 (.23) [.35 - .43]
Formal (F)	184	57.3	.52	.49 (.19) [.46 - .52]
Predictor Timing Format				
Cannot Tell	39	12.1	.54	.53 (.12) [.37 - .69]
Latency	36	11.2	.39	.33 (.26) [.22 - .44]
Continuous	246	76.7	.47	.48 (.19) [.46 - .50]

Criterion Norm				
National	82	25.5	.45	.44 (.20) [.40 - .48]
State	239	74.5	.45	.51 (.22) [.47 - .54]
Ethnicity				
Caucasian	213	66.4	.45	.49 (.23) [.46 - .52]
African American	34	10.5	.32	.35 (.19) [.29 - .47]
Mixed Sample	74	23.1	.51	.51 (.16) [.47 - .55]
Gender				
Cannot Tell	141	43.9	.46	.39 (.23) [.35 - . 43]
Male	104	32.4	.49	.42 (.19) [.38 - .46]
Female	76	23.7	.61	.61 (.13) [.58 - .64]
Socioeconomic Status				
Cannot Tell	154	48.0	.49	.45 (.17) [.42 - .48]
Free & Reduced Lunch	44	13.7	.38	.43 (.14) [.39 - .47]
Full Priced Lunch	123	38.3	.54	.47 (.28) [.42 - .52]
Education Status				
Cannot Tell	75	23.4	.54	.52 (.12) [.47 - .57]
General Education	246	76.6	.46	.44 (.23) [.39 - .48]
Location				
Cannot Tell	16	4.9	.48	.48 (.09) [.44 - .52]
Midwest	44	13.7	.45	.46 (.11) [.43 - .49]
Northeast	110	34.3	.51	.45 (.27) [.40 - .50]
Northwest	92	28.6	.52	.47 (.16) [.44 - .51]
Southeast	59	18.4	.49	.44 (.17) [.40 - .48]
Sample Type				
Cannot Tell	240	74.8	.50	.45 (.24) [.42 - .48]
Random	47	14.6	.46	.45 (.15) [.41 - .49]
Convenience	34	10.6	.46	.47 (.11) [.43 - .51]

Examiner Qualifications				
Cannot Tell				
Pre-Professional	3	.9	.39	.39 (.16) [.31 - .57]
Professional	94	29.8	.44	.38 (.28) [.32 - .44]
Other	77	24.0	.58	.56 (.16) [.42 - .60]
Mixed	77	24.0	.49	.49 (.20) [.44 - .53]
	70	21.8	.53	.51 (.12) [.43 - .59]

Figure 1 is a representation of the percentage of participants in each of the grade levels in the total sample across all sixteen inclusive studies. Grade level ranged from Pre-K to second grade and included a category for samples that were composed of participants across grade levels. This figure indicates that the majority of samples were made up of kindergarten and first grade students. *Figure 2* represents the distribution of the skill category that the predictors measured. There were a total of 48 early numeracy measures with different names which constitute five skill categories. This suggests a considerable amount of overlap between names of EM-CBM measures. Similarly, *figure 3* depicts the representation of the 29 different criterion measures and the six categories that define the skill the measures purport to assess.

Moderator Analysis

To examine the influence of potential moderating variables, each variable was separated into categories. For each specific group, the correlations were corrected for sampling error and the mean correlation was computed as previously described [(r * n) / *N*]. The confidence intervals around the uncorrected correlations were calculated. Within each group condition, it was examined whether the coded variables significantly departed from each other. A significant deviation was defined as non overlapping confidence intervals between each of the variables.

The following variables were then further analyzed using a one-way analysis of variance: correlation type, predictor measure skill category, criterion measure skill category, grade level of participants, examiner training, methods taken to ensure procedural integrity, and the category of the predictor measure. A one-way analysis of variance (ANOVA) was conducted to evaluate the relationships within each coded variable. Table 3 provides the results of the ANOVA and post hoc tests comparing the potential moderating variables.

Table 3. *Statistical Results of Moderating Variables*

Moderator	Mean	*F* score	*p* value	Effect Size	Post Hoc Results	Significance Level
Correlation Type		7.34	.01	.02		
Concurrent (CN)	.48				CN < PR	.01
Predictive (PR)	.51					
Predictor Skill		2.85	.02			
Counting (CN)	.46				RC < CN	.02
Comparing and Ordering (CO)	.57				RC < CO	.01
Adding to/Taking Away (AT)	.51				RC < AT	.01
Mixed Skill (MS)	.55				RC < MS	.02
Readiness Concepts (RC)	.06				CN < CO	.01
					CN < MS	.02
Criterion Skill		1.97	.05	.07		
Cannot tell (CT)	.47				AP < MS	.02
Basic Readiness Concepts (BR)	.51				AP < BR	.04
Basic Math Concepts (BM)	.45				BM < MS	.01
Applications/Problem Solving (AP)	.41				BM < BR	.03
Calculation of Non-basic Facts (CF)	.54					
Mixed Skill (MS)	.53					
Grade Level		2.74	.05	.07		
Pre-Kindergarten (PK)	.41				PK < S	.04
Kindergarten (K)	.50				PK < M	.04
First (F)	.44				F < S	.02
Second (S)	.56					

Mixed (M)	.54					
Examiner Training		0.90	.30	.00		
Cannot tell (CT)	.48		.			
Trained (T)	.49					

Moderator	Mean	*F* score	*p* value	Effect Size	Post Hoc Results	Significance Level
Procedural Integrity		2.90	.02	.18		
Cannot tell (CT)	.34				CT < UD	.00
Used Data (UD)	.54				CT < ND	.03
No Data (ND)	.49				CT < N	.02
Narrative (N)	.52					
Predictor Category		2.35	.02	.07		
Cannot tell (CT)	.70				I < CT	.01
Informal (I)	.45				F < CT	.01
Formal (F)	.52				I < F	.04

Correlation type. A one-way analysis of variance was conducted to evaluate the relationship between correlation types. The independent variable, correlation type, included two levels: predictive and concurrent. The dependent variable was the correlation between EM-CBM probes and math achievement assessments. The ANOVA was significant, $F(1,72) = 7.34$, $p = .01$. The strength of relationship between correlation type and the relationship between EM-CBM and standardized math achievement assessments, as assessed by $\acute{\eta}^2$, was weak, and indicates that correlation type accounts for 2% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the two groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the predictive validity correlations and the concurrent validity correlations. Predictive validity correlations showed a stronger relationship than concurrent validity correlations. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Predictor skill category. A one-way analysis of variance was conducted to evaluate the relationship between predictive skill categories. The independent variable, predictor skill category, included five groups: counting, comparing and ordering, adding to/taking away, mixed skill, and readiness concepts. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was significant, $F(4, 316) = 2.85$, $p = .02$. The strength of relationship between predictor skill category and the relationship between EM-CBM and standardized math achievement assessments, as assessed by $\acute{\eta}^2$, was weak, and indicates that predictor skill category accounts for 7% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the five groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the readiness concepts and counting, comparing and ordering, adding to/taking away, and mixed skill probes. Readiness concept probes showed a weaker relationship than the aforementioned skill probes. There was a significant difference in the means between the counting and comparing and ordering and mixed skill probes. Probes that measured counting ability demonstrated a weaker relationship than comparing and ordering and mixed skill probes. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Criterion skill category. A one-way analysis of variance was conducted to evaluate the relationship between criterion skill categories. The independent variable, criterion skill category, included six groups: cannot tell, basic readiness concepts, basic math concepts, applications/problem solving, calculation of non-basic facts, and mixed skill. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was significant, $F(5, 312) = 1.97$, $p = .05$. The strength of relationship between predictor skill category and the relationship between EM-CBM and standardized math achievement assessments, as assessed by $\acute{\eta}^2$, was weak, and indicates that criterion skill category accounts for 6% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the six groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the applications/problem solving and mixed skill assessments, applications/problem solving and basic readiness concepts assessments, basic math concepts and mixed skill assessments, and basic math concepts and basic readiness concepts assessments. Applications/problem solving assessments showed a weaker relationship than mixed skill and basic readiness assessments. Basic math concepts assessments demonstrated a weaker relationship than mixed skill and basic readiness assessments. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Grade level. A one-way analysis of variance was conducted to evaluate the relationship between grade level and its relationship with early numeracy and math achievement. The independent variable, grade level, included five groups: pre-kindergarten, kindergarten, first grade, second grade, and mixed grade level. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was significant, $F(4, 316) = 2.74$, $p = .05$. The strength of relationship between grade level and the relationship between EM-CBM and standardized math achievement assessments, as assessed by ή², was weak, and indicates that grade level accounts for 6% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the five groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the pre-kindergarten and second grade, pre-kindergarten and mixed grade level, and first grade and second grade. Second grade probes showed a stronger relationship than the pre-kindergarten probes and the first grade probes. Mixed grade level probes also demonstrated a stronger relationship than the pre-kindergarten probes. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Examiner training. A one-way analysis of variance was conducted to evaluate the relationship between levels of examiner training. The independent variable, examiner training, included three possible groups: cannot tell, trained, and untrained. However, within the examined articles, there were no reported untrained examiners. For this reason, examiner training only contained two separate groups: cannot tell and trained. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was not significant, $F(1, 319) = 0.9$, $p = .30$. The strength of relationship between examiner training and the relationship between EM-CBM and standardized math achievement assessments, as assessed by ή², was very weak, and indicates that level of examiner training accounts for 0% of the variance of the dependent variable. The ANOVA was not significant, therefore there were no follow-up tests conducted. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Procedural integrity. A one-way analysis of variance was conducted to evaluate the relationship between measures of procedural integrity. The independent variable, procedural integrity, included four groups: cannot tell, used data, no data, and narrative. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was significant, $F(3, 317) = 2.90$, $p = .02$. The strength of relationship between procedural integrity and the relationship between EM-CBM and standardized math achievement assessments, as assessed by ή², was moderate, and indicates that procedural integrity accounts for 18% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the four groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the cannot tell group and each of the other three groups: used data, no data, and narrative. Studies that used procedural integrity data, used no data, and presented narrative procedural integrity data all demonstrated a stronger relationship with the correlation between early numeracy and math achievement than studies that were coded as cannot tell. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

Predictor category. A one-way analysis of variance was conducted to evaluate the relationship between predictor categories. The independent variable, predictor category, included three groups: cannot tell, informal, and formal. The dependent variable was the correlation between early numeracy and math achievement. The ANOVA was significant, $F(2, 318) = 2.35$, $p = .02$. The strength of relationship between predictor category and the relationship between EM-CBM and standardized math achievement assessments, as assessed by $\acute{\eta}^2$, was weak, and indicates that predictor category accounts for 7% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pariwise differences among the means. Because the variances among the three groups were not equal, it was not assumed that the variances were homogeneous and conducted post hoc comparisons with the use of the Dunnett's C test, a test that does not assume equal variances among the groups. There was a significant difference in the means between the informal probes and formal probes, informal probes and cannot tell, and formal probes and cannot tell. Informal probes showed a weaker relationship than the formal probes. Probes that coded as cannot tell demonstrated a stronger relationship than both informal and formal probes. The means and standard deviations of the groups are reported in Table 2. The results of the ANOVA tests are reported in Table 3.

CHAPTER V: DISCUSSION

Relationship between Early Numeracy and Math Achievement

The primary focus of the present study was to summarize the literature that examined the criterion-related validity of curriculum-based measures of early numeracy. The main research objective was an investigation into the strength and direction of the correlational relationship between early numeracy measures and math achievement assessment tools. There were 16 studies analyzed, and 321 correlations were corrected for sampling error and then averaged to obtain an overall correlation coefficient.

The first hypothesis was that there would be a strong, positive correlation between the predictor and criterion measures. The results support this hypothesis and indicate that the mean correlation between early numeracy and math achievement is .49. This correlation coefficient signifies a moderate-to-strong relationship between the two variables. This suggests that the higher a child scores on a single measure of early math numeracy, the higher he or she would score on a standardized, norm-referenced measure of math achievement. Furthermore, measuring one specific subskill of early numeracy, such as comparing and ordering, can be an indicator of a child's overall level of math achievement. These results are consistent with the literature base examining the concurrent and predictive validity that exists on measures of early numeracy for children in Pre-Kindergarten through second grade.

Moderator Analyses

The second objective of this study was to examine the variables which influence the relationship between early numeracy and math achievement and determine which variables are moderators. The second hypothesis of the study was that the overall correlation will be affected by hypothetical moderators, which may include one or a combination of numerous variables such as grade level, gender, location, ethnicity, SES, education status, and administration format. Results of the one-way analyses of variance were not in full support of the hypothesis. There were six variables that were identified as moderators, which affected the overall correlation coefficient; however, these variables were other than those established a priori. The variables that were found to moderate the correlational relationship are the following: correlation type, predictor skill, criterion skill, grade level, procedural integrity, and predictor category. Specifically, these six variables were qualitative variables found to influence the strength of the relationship between the predictor and criterion variables (Baron & Kenny, 1986).

This suggests that it is necessary to take into consideration these factors when determining the relationship between a measure of early numeracy and overall level of math achievement. For example, a child in the second grade is expected to have a stronger correlation between the two variables than a child in Pre-Kindergarten. Meaning, the score of a child's EM-CBM probe at the second grade level will be better able to predict the same child's score on a standardized, norm-referenced achievement test such as the Woodcock-Johnson Tests of Achievement, III.

Correlation type. It is evident that the type of correlation is a variable that moderates the relationship between early numeracy and math achievement. This suggests EM-CBM probes that are administered earlier than at least one month prior to math achievement assessments have a greater relationship than early numeracy probes that are administered within one month of the math achievement assessments. For this reason, when a practitioner is using an early numeracy probe to predict math achievement, it is important to administer the early numeracy measure greater than one month prior to the standardized achievement measure.

Predictor skill. It is evident that predictor skill category is a variable that moderates the relationship between early numeracy and math achievement. The mean correlations of several of the different skill categories were significantly different. In terms of predictor skill, EM-CBM measures that specifically assess comparing and ordering have the strongest correlation with math achievement tests (.57), while EM-CBM measures of readiness concepts have the weakest correlation (.06). This suggests that measures of readiness concepts have poor utility when used to predict math achievement. Specifically, EM-CBM readiness concepts measures were different than all other early numeracy measures. EM-CBM counting measures are significantly different than EM-CBM comparing and ordering measures and mixed skill measures. Therefore, the relationship between early numeracy measures and math achievement is moderated by the skill category of the early numeracy measure.

Criterion skill. Criterion skill category is also a moderator between the relationship of early numeracy and math achievement. Standardized math achievement tests that involved calculation tasks of non basic facts were found to have the highest correlation with the predictor measures (.54) followed by mixed skill assessments (.53). Criterion measures that utilized applications/problem solving methods were found to have the weakest correlation with the predictor measures, but it was still a moderate relationship (.41). This may be due to the fact that few, if any, of the early numeracy probes measured any application type problems. Achievement tests that measure applications/problem solving have a significantly different mean correlation than basic readiness concepts and mixed skills assessments. Achievement tests that measure basic math concepts have a significantly different mean correlation than tests that measure mixed skills and basic readiness concepts.

Grade level. The grade level of the student being assessed moderates the relationship between early numeracy and math achievement. Children obtained overall correlations significantly different depending on their grade level. Specifically, second grade students were found to be significantly different than pre-kindergarten students and students in first grade. Pre-kindergarten students also obtained significantly different mean scores than students in a mixed grade level. This suggests that as the grade level of a student changes, the relationship between early numeracy and math achievement will vary. Children in the second grade demonstrated the strongest correlation (.56) followed by mixed grade level (.54), kindergarten (.50), first grade (.44), and Pre-Kindergarten (.41). This may be associated with the fact that as children progress developmentally, they are better able to learn math concepts. The greater range of skills one can assess, the more utility a practitioner gains by examining a variety of strengths and weaknesses of a student.

Procedural integrity. It was recorded whether articles used data or narrative to present the measures taken to ensure procedural integrity; or if there was no data or it was unable to determine whether procedural measures were taken. After conducting a one-way analysis of variance and post hoc tests, it is evident that articles that were coded “cannot tell” were significantly different than articles that presented data and articles that used narrative form. This suggests that procedural integrity may be a moderating variable. Articles that used data to present their methods of procedural integrity obtained the strongest correlations (.54) followed by articles that used narrative form (.52). Articles that did not specifically indicate any procedural integrity methods were found to have the weakest correlations between the predictor and criterion variables (.34). This suggests that it is important for researchers to take precautions to ensure that the data being collected and entered are accurate. The more confident a practitioner is in the accuracy of the data, the stronger the relationship between measures of early numeracy and math achievement. For example, if procedures have been taken to ensure the integrity of the dataset, then increased utility can be gained from an EM-CBM probe in its ability to predict the overall level of math achievement. However, it should be noted that while these results suggest that procedural integrity may be a moderating variable, there were no significant differences between those articles that presented methods of procedural integrity in varying ways. This indicates that these results must be interpreted with caution and further analysis should be conducted to determine whether procedural integrity is a true moderator.

Predictor category. Both informal and formal measures of EM-CBM probes exist. It was recorded whether EM-CBM measures were informal, formal, or unable to determine. The predictor category was found to be a moderator in this meta-analysis. Informal measures of early numeracy and formal measures of early numeracy were significantly different than measures that it was unable to determine. Informal and formal measures were also found to have significantly different means. This suggests that the correlation between early numeracy and math achievement varies depending on the category of the predictor measure. EM-CBM probes that were not identified as either formal or informal obtained the strongest correlation coefficient in this analysis (.70). However, these results are not suggestive that predictor category is unimportant. These results must be interpreted with caution due to the fact that there was only one EM-CBM measure (Number Sense) that was coded “cannot tell” which accounted for 6 total correlations. It may be concluded that it is important for EM-CBM probes to identify whether it is an informal or formal measure of mathematics. The Number Sense EM-CBM measure was coded cannot tell because it was assessing both formal and informal tasks. It should also be noted that formal measures of early numeracy have a stronger relationship (.52) with math achievement than informal measures (.45). Possibly, this is because the majority of criterion assessments involve the use of written numerals.

Limitations and Directions for Future Research

A limitation of this study is that an article that pertains to this topic may have been unintentionally missed, or published after the data collection, and therefore not included in the analysis. Although efforts were taken to collect each relevant article, there were a few articles that were not accessible to the author in order to determine whether it met inclusion criteria. Correcting for sampling error was an attempt to account for difference in sample size; however, it still remains a threat to the design of a meta-analysis. Another limitation of the present study is that research was conducted on varying participants in the inclusive studies. Including participants from a variety of locations, socio-economic backgrounds, ethnicities, and education levels allows for greater generalization across studies. However, this serves as a threat to validity when comparing the different students against each other.

It would be beneficial for future research to take the moderating variables and enter them into a hierarchical linear regression analysis in order to determine the amount of variance each variable can account for in the relationship between early numeracy and math achievement. The present study has identified which variables are moderators; however, a further analysis would also identify which moderators have a greater influence on the overall correlational relationship.

Implications

Due to the lack of a current research synthesis, conducting a meta-analysis on this topic allowed the organization and quantification of research studies which examined the correlation of EM-CBM and math criterion measures. The present study sought to quantify an overall correlation and a correlation that takes into account the following variables: sample size, grade level, gender, location, ethnicity, SES, education status, and administration format.

A significant outcome of this study are the results of the meta-analysis itself. As a moderate-to-strong positive correlation exists, researchers and practitioners can become more confident in the utility of EM-CBM. That is, with increased confidence in the predictive validity of EM-CBM, a norm-referenced assessment will not always need to be administered to assess a child’s academic abilities. For those EM-CBM measures which are determined to be valid predictive measures, practitioners will be able to screen and identify children in need of interventions using only an EM-CBM measure, which will be time and cost efficient.

The measures which are most valid are those that obtained the largest correlations with criterion measures. This indicates that the measures that obtained the strongest relationship with math achievement measures are measuring the skills that are intended to be measured. The early numeracy measures that demonstrated the greatest relationship with the criterion measures are the following: comparing and ordering probes, mixed skill probes, adding to/taking away probes, and counting measures. This suggests that the aforementioned early numeracy measures are best able to predict a child’s score on a standardized math achievement assessment. Furthermore, it can be determined that these early numeracy measures are a greater predictor of math achievement than the other predictor category skills. This is important in a practitioner setting so that student’s can be assessed using the early numeracy measures with the greatest validity and strongest relationship. However, standardized tests will still be necessary to determine eligibility for specific programs or to diagnose learning disabilities and gifted children. Finally, it is important to take into consideration the variables that moderate the predictor-criterion relationship when using either or both of the measures of mathematics.

The EM-CBM skill categories that obtained weak correlations with math achievement, such as counting, are still important skills to assess. Caution needs to be used when interpreting and generalizing a child's score on an EM-CBM counting probe. This further suggests that practitioners should select a different skill probe to use when attempting to predict a child's overall math achievement ability. However, an EM-CBM counting probe is still an effective way to assess a child's ability informal, conceptual knowledge of numbers and counting. Oftentimes, before assessing a child on a more challenging skill, it is first necessary to ensure mastery level on the easier skills. For these reasons, it is important to know the relationship between early numeracy and math achievement in order to gain insight into the function of each EM-CBM probe. It is also imperative for practitioners to understand the utility of each measure. Specifically, it is important to understand whether a measure should be used for its predictive ability or assessment purposes.

Conclusions

Early numeracy is an important construct to assess when identifying a child's level of achievement. Measuring early numeracy can a) help to identify those students who may be at risk for academic failure, and b) as monitor the progress of students as they make performance gains. Many EM-CBM measures are designed to assess specific subskills which allow practitioners to identify the skills that students need assistance with.

Measuring early numeracy is also helpful in predicting a child's level of overall math achievement. Math achievement is typically measured using standardized, norm-referenced assessment tools. The results of this meta-analysis conclude that early numeracy measures and standardized math achievement assessments are moderately-to-strongly correlated in a positive direction. This suggests that practitioners can be confident in the utility of the EM-CBM measures when predicting a child's math achievement level.

However, variables have been identified that moderate the strength and direction of the predictor-criterion relationship. In practice, when practitioners are using the early numeracy measures as a predictor of math achievement level, these moderating variables need to be taken into consideration before reaching any conclusions. Specifically, a practitioner must consider the skill category of both the predictor and criterion measures as well as the grade level of a student before determining the accurate correlational relationship specific to that individual child. Overall, it may be concluded that the EM-CBM measures are moderate-to-strong predictors of a child's math achievement level.

REFERENCES

Aunola, K., Leskinen, E., Lerkkanen, M.K., & Nurmi, J.E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96(4),* 699-713.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Baroody, A.J. (2004). The developmental bases for early childhood operations and number standards. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education (*pp. 173-220). Mahwah, NJ: Erlbaum.

Bear, G.G., Minke, K.M., & Manning, M.A. (2002). Self-concept of students with learning disabilities: A meta-analysis. *School Psychology Review, 31(3),* 405-427.

Berch, D.B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 35(4),* 333-339.

Bracken, B.A. (2006). Bracken basic concepts scale, third edition. Birmingham, AL: PsychCorp.

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

*Clarke, B., Baker, S., Smolkowski, K., & Chard, D.J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29(1),* 46-57.

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

Clements, D.H. (2004). Major themes and recommendations. In D.H. Clements, J. Sarama, & M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 7-76). Mahwah, NJ: Erlbaum.

Clements, D.H., Sarama, J., & DiBiase, M. (Eds.). (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education.* Mahwah, NJ: Erlbaum.

*Connell, J.E. (2005). Constructing a math applications, curriculum-based assessment: An analysis of the relationship between applications problems, computation problems, and criterion-referenced assessments (Doctoral dissertation, Louisiana State University, 2005).

Cooper, H. (1998). *Synthesizing research, Third Edition: A guide for literature reviews.* Thousand Oaks, CA: Sage.

Cooper, H., Appelbaum, M., Maxwell, S., Stone, A., & Sher, K.J. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63(9))*, 839-851.

Cooper, H., & Hedges, L.V. (Eds.). (1994). Using oral reading rate to predict student performance on statewide assessment tests. *The handbook of research synthesis.* New York: Russell Sage Foundation.

Crawford, L., Tindal, G., & Steiber, S. (2001). *Educational Assessment, 7(4)*, 303-323.

Daly, E.J., Lentz, F.E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11(4),* 369-386.

*Daly, E.J., Wright, J.A., Kelly, S.Q., & Martens, B.K., (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12(3),* 268-280.

Delazer, M., Girelli, L., Grana, A., & Domahs, F. (2003). Number processing and calculation: Normative data from healthy adults. *The Clinical Neuropsychologist, 17,* 331-350.

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

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

Eccles, J. (1997). User-friendly science and mathematics: Can it interest girls and minorities in breaking through the middle school wall? In D. Johnson et al. (Eds.). *Minorities and girls in school: Effects on achievement and performance.* Vol. 1, p. 65-104. Thousand Oaks, CA: Sage.

*Floyd, R.G., Hojnoski, R., & Key, J. (2006). Preliminary evidence of the technical adequacy of the preschool numeracy indicators. *School Psychology Review, 35(4),* 627-644.

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

Fuchs, L.S., & Deno, S.L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488-500.

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

Fuchs, L.S., & Fuchs, D. (1986). Effects of systematic formative evaluation on student achievement. *Exceptional Children, 53,* 199-207.

Fuchs, L.S., & Fuchs, D. (1988). Curriculum-based measurement: A methodology for evaluating and improving student programs. *Diagnostique, 14(1),* 3-13.

Fuchs, L.S., Fuchs, D., Bishop, N., & Hamlett, C.L. (1992). Classwide decision-making strategies with curriculum-based measurement. *Diagnostique, 18(1),* 39-52.

*Fuchs, L.S., Fuchs, D., Compton, D.L., Bryant, J.B., 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.

Geary, D.C. (1996). International differences in mathematical achievement: Their nature, causes, and consequences. *Current Directions in Psychological Science, 5,* 133-137.

Ginsburg, H.P., & Baroody, A.J. (2003). *Test of Early Mathematics Ability, Third Edition,* Austin, TX: Pro-Ed.Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Charles E. Merrill Publishing Co.

Hintze, J.M., Christ, T.J., & Methe, S.A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43(1),* 45-56.

Hartman, J.M., & Fuller, M.L. (1997). The development of curriculum-based measurement norms in literature-based classrooms. *Journal of School Psychology, 35,* 377-389.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

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

Hunter, J.E. & Schmidt, F.L. (2004). Methods of meta analysis: Correcting error and bias in research findings. Thousand Oaks: Sage.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology, 91*, 116-126.

*Jordan, N.C., Kaplan, D., Locuniak, M.N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research and Practice, 23(1),* 36-46.

*Lago, R.M. (2007). Examining the psychometrics of number sense among kindergarten students (Doctoral dissertation, The Pennsylvania State University, 2007).

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

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

*Martinez, R.S., Missall, K.N., Graney, S.B., Aricak, O.T., & Clarke, B. (2008). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention, 34, 2*, 116-125.

Mazzocco, M.M., & Thompson, R.E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research and Practice, 20(3),* 142-155.

McBee, M.T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education, 17(2),* 103-111.

*Methe, S.A., Hintze, J.M., & Floyd, R.G. (2008). Validation and decision accuracy of early numeracy skill indicators. *School Psychology Review, 37(3),* 359-373.

Methe, S.A., & Riley-Tillman, T.C. (2008). An informed approach to selecting and designing early mathematics interventions. *School Psychology Forum: Research in Practice, 2(3),* 29-41.

Mirkin, P.K., Deno, S.L., Tindal, G., & Kuehnle, K. (1982). Frequency of measurement and data utilization strategies as factors in standardized behavioral assessment of academic skill. *Journal of Behavioral Assessment, 4(4),* 361-370.

Naglieri, J.A., & Crockett, D.P. (2005). Response to Intervention (RTI): Is it a scientifically proven method? *Communique,* 34.

National Center for Education Statistics. (2009). *National Assessment of Educational Progress: The Nation's Report Card: Mathematics 2007.* Washington, DC: United States Department of Education- Institute of Education Sciences.

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

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics.* London: Blackwell.

Nunnally, J.C., & Bernstein, I.H. (1995). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61,* 27-59.

Reid, E.E., Morgan, P.L., DiPerna, J.C., & Lei, P.W. (2006). Development of measures to assess young children's early academic skills: Preliminary findings from a Head Start-university partnership. *Insights on Learning Disabilities, 3,* 25-38.

Riley, R.W. (1997). *Mathematics equals opportunity*. District of Columbia, U.S.: Federal Department of Education.

Rivera-Batiz, F.L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *The Journal of Human Resources, 27,* 313-328.

Rohling, M.L., Faust, M.E., Beverly, B., & Demakis, G. (2009). Effectiveness of cognitive rehabilitation following acquired brain injury: A meta-analytic re-examination of Cicerone et al.'s (2000, 2005) systematic reviews. *Neuropsychology, 23(1),* 20-39.

Russell, R.L., & Ginsburg, H.P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction, 1,* 217-244.

Shaughnessy, J.J., Zechmeister, E.B., & Zechmeister, J.S. (2008). *Research methods in psychology*. New York: McGraw-Hill Humanities Social

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

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

Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society,* Ser. B. 13, 328-241.

Stecker, P.M., & Fuchs, L.S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15(3),* 128-134.

Stevenson, H.W., & Newman, R.S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646-659.

Swanson, H.L., Trainin, G., Necoechea, D.M., & Hammill, D.D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73(4),* 407-440.

*VanDerHeyden, A.M., Broussard, C., & Cooley, A. (2006). Further development of measures of early math performance for preschoolers. *Journal of School Psychology, 44,* 533-553.

*VanDerHeyden, A.M., Broussard, C., Fabre, M., Stanley, J., Legendre, J., & Creppell, R. (2004). Development and validation of curriculum-based measures of math performance for preschool children. *Journal of Early Intervention, 27(1),* 27-41.

*VanDerHeyden, A.M., Witt, J.C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30(3),* 363-382.

Woolfolk, A. (2009). *Educational Psychology*, eleventh edition. New Jersey: Prentice Hall.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). Woodcock-Johnson tests of achievement, third edition. Rolling Meadows, IL: Riverside Publishing.

Yoshida, H., & Sawano, K. (2002). Overcoming cognitive obstacles in learning fractions: Equal-partioning and equal-whole. *Japanese Psychological Research, 44(4),* 183-195.

Table 1. *Technical Adequacy Data for Early Numeracy Curriculum-Based Measures*

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
Floyd, Hojnoski, & Key (2006)	Mixed	41	• One-to-One Correspondence Counting Fluency (OOCCF) • Oral Counting Fluency (OCF) • Number Naming Fluency (NNF) • Quantity Comparison Fluency (QCF)	Test-retest: OOCCF = .62, .96 OCF = .90, .83 NNF = .91, .88 QCF = .89, .94	C: BBCS-R OOCCF = .41 OCF = .57 NNF = .43 QCF = .61 C: BBCS-Q OOCCF = .36 OCF = .36 NNF = .34 QCF = .52 WJ-III (AP) OOCCF = .40 OCF = .45 NNF = .49 QCF = .51 TEMA-3 OOCCF = .54 OCF = .55 NNF = .60 QCF = .48
Fuchs, Fuchs, Compton, Bryant, Hamlett, & Seethaler (2007)	1	170	• Fact Retrieval (FR) • Number Identification/Counting (NIDC)	Internal Consistency: FR = .84 NIDC = .92 .	C: MCBM FR = .34 C: CBM-C/A FR = .15 NIDC = .36 P: WRAT-(ARITH) FR = .14 NIDC = .34

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					P: WP FR = .10 NIDC = .39
Clarke, Baker, Smolkowski, & Chard (2008)	K	221	• Oral Counting (OC) • Number Identification (NI) • Missing Number (MN) • Quantity Discrimination (QD)	__	C: SESAT OC = .59 NI = .53 MN = .60 QD = .62 OC = .55 NI = .61 MN = .64 QD = .62 P: SESAT OC = .55 NI = .58 MN = .57 QD = .60
Lembke & Foegen (2009)	K	44-88	• Quantity Discrimination (QD) • Quantity Array (QA) • Number Identification (NI) • Missing Number (MN)	Alternate form: QD = .89, .88, .83 QA = .84, .74, .81 NI = .91, .92 MN = .72, .75, .59 Test-retest: QD = .85 QA = .72, .80 NI = .88 MN = .80, .84	C: MBA QD = .50 QA = .38 NI = .52 MN = .57 QD = .38 QA = .49 MN = .49 TEMA-3 QD = .45

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					QA = .29 NI = .33 MN = .48 P: TEMA-3 QD = .35 QA = .35 NI = .34 MN = .37 C: MBA QD = .31 QA = .43 NI = .49 MN = .44 QD = .48 QA = .37 MN = .45 SESAT: QD = .60 QA = .57 NI = .52 MN = .75 TEMA-3 QD = .57 QA = .60 NI = .52 MN = .54
	1	28-126	• Quantity Discrimination (QD) • Quantity Array (QA) • Number Identification (NI) • Missing Number (MN)	Alternate form: QD = .86, .89, .81 QA = .85, .83, .80 NI = .87, .90 MN = .73, .80, .81 Test-retest: QD = .83, .91 QA = .86, .85 NI = .88 MN = .81, .88	

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					P: TEMA-3 QD = .43 QA = .51 NI = .58 MN = .68
Methe, Hintze, & Floyd (2008)	K	64-77	• Counting-on Fluency (COF) • Match Quantity Fluency (MQF) • Number Recognition Fluency (NRF) • Ordinal Position Fluency (OPF)	Test-retest: COF = .68 NRF = .98 MQF = .74 OPF = .81 Internal consistency: COF = .80 MQF = .53 OPF = .83	C: TEMA-3 COF = .50 MQF = .72 NRF = .55 OPF = .63 P: TEMA-3 COF = .55 MQF = .64 NRF = .20 OPF = .60 COF = .46 MQF = .70 NRF = .41 OPF = .58 COF = .62 MQF = .66 NRF = .47 OPF = .57
Lembke, Foegen, Whittaker, & Hampton (2008)	1	30	• Quantity Discrimination (QD) • Number Identification (NI) • Missing Number (MN)	Alternate-form: QD = .80 NI = .77 MN = .79	C: SESAT QD = .50 NI = .47 MN = .21

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
Martinez, Missall, Graney, Aricak, & Clarke (2008)	K	52	• Oral Counting (OC) • Number Identification (NI) • Quantity Discrimination (QD) • Missing Number (MN)	Test-retest: NI = .92 QD = .80 MN = .89 Alternate-form: NI = .91 QD = .77 MN = .79	P: SAT-10 OC = .45 NI = .31 QD = .46 MN =. 36 C: SAT-10 NI = .44 QD = .63 MN = .47
VanDerHeyden, Broussard, & Cooley (2006)	Mixed	38-42	• Count Objects (CO) • Choose Number (CN) • Discrimination (D0 • Number Naming (NN) • Free Count (FC)	Alternate-form: CO = .83 CN = .87 D = .88 FC = .71	P: CBM-CN CO = .50 CN = .31 D = .60 NN = .46 FC = .57
Daly, Wright, Kelly, & Martens (1997)	1	*30*	• Color Naming (CN) • Number Reading (NR) • Number Counting (NC) • Number Production (NP) • Number Selection (NS) • Shape Naming (SN)	Test-retest: CN = .78 NR = .82 NC = .88 NP = .37 NS = .67 SN = .47	C: WJ-R-BM CN = .05 NR = .03 NC = .47 NP = .17 NS = .11 SN = .09 P: CBM-ADD CN = .06

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					NR = .07 NC = .39 NP = .36 NS = .30 SN = .04 CBM-SUB CN = .06 NR = .07 NC = .39 NP = .36 NS = .30 SN = .04
Connell (2005)	1	39	• Single Skill Computation (SSC) • Multiple Skill Computation (MSC)	___	C: WJ-III BS SSC = .70 MSC = .76 WJ-III CALC SSC = .46 MSC = .40 WJ-III FL SSC = .71 MSC = .66 WJ-III AP SSC = .64 MSC = .78

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
	2	28	• Single Skill Computation (SSC) • Multiple Skill Computation (MSC)	___	SSC = .43 MSC = .66 WJ-III CALC SSC = .45 MSC = .62 WJ-III FL SSC = .78 MSC = .74 WJ-III AP SSC = .25 MSC = .52
Lago (2007)	K	20	• Counting Aloud (CA) • Counting Objects (CO) • Measurement concepts (MC) • Nonverbal Calculation (NC) • Number Identification (NID) • Quantity Discrimination (QD) • Estimation (EST) • Rapid Naming Objects (RND) • Rapid Naming Colors(RNC) • Rapid Naming Numbers(RNN)	Test-retest: CA = .79 CO = .60 MC = .85 NC = .81 NID = .62 QD = .93 EST = .88 RND = .58 RNC = .90 RNN = .51 Internal consistency: CO = .73 MC = .61 NID = .66 EST = .91	C: WJ-III CALC CA = .52 CO = .30 MC = .33 NC = .26 NID = .56 QD = .35 EST = .17 RND = -.37 RNC = -.21 RNN =-.10 WJ-III AP CA = .70 CO = .15 MC = .21 NC = .62

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
				RND = .86 RNN =.61 Alternate form: CA = .93 NID = .89, .93 QD = .85, .93 RND = .78 RNC = .78 RNN =.78	NID = .50 QD = .42 EST = .30 RND = -.38 RNC = -.22 RNN = -.12 WJ-III QC CA = .38 CO = -.01 MC = .01 NC = .24 NID = .47 QD = .23 EST = .11 RND = -.06 RNC = .00 RNN = -.03 WJ-III MR CA = .58 CO = .10 MC = .12 NC = .48 NID = .54 QD = .36 EST = .25 RND = -.27 RNC = -.13

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					RNN = -.09
Jordan, Kaplan, Locuniak, & Ramineni (2007)	Mixed	277	• Number Sense (NS) • Counting Skills (CS) • Number Knowledge (NK) • Nonverbal Calculation (NC) • Story Problems (SP) • Number Combinations (NCO)	___	WJ-III MATH NS = .70 CS = .36 NK = .54 NC = .52 SP = .47 NCO = .58 NS = .66 CS = .37 NK = .57 NC = .40 SP = .52 NCO = .49 NS = .69 CS = .36 NK = .52 NC = .53 SP = .54 P: WJ-III MATH NCO = .58 NS = .73 CS = .35 NK = .59 NC = .58 SP = .62

Author(s)	Grade	*N*	Early Numeracy Curriculum-Based Measure	EM-CBM Reliability	Criterion validity[a]
					NCO = .64 NS = .71 CS = .30 NK = .53 NC = .50 SP = .59 NCO = .65 NS = .72 CS = .28 NK = .54 NC = .51 SP = .62 NCO = .68

Note. See Foegen, Jiban, and Deno (2007) for a more comprehensive review of technical adequacy of EM-CBM studies. Refer to Appendix A for a complete listing of EM-CBM abbreviations.

[a]P = predictive criterion validity data; C = concurrent criterion validity data.

Figure Captions

Figure 1. Grade level composition of overall sample

Figure 2. Predictor skill categories included in the meta-analysis

Figure 3. Criterion skill categories included in the meta-analysis

Figure 1.

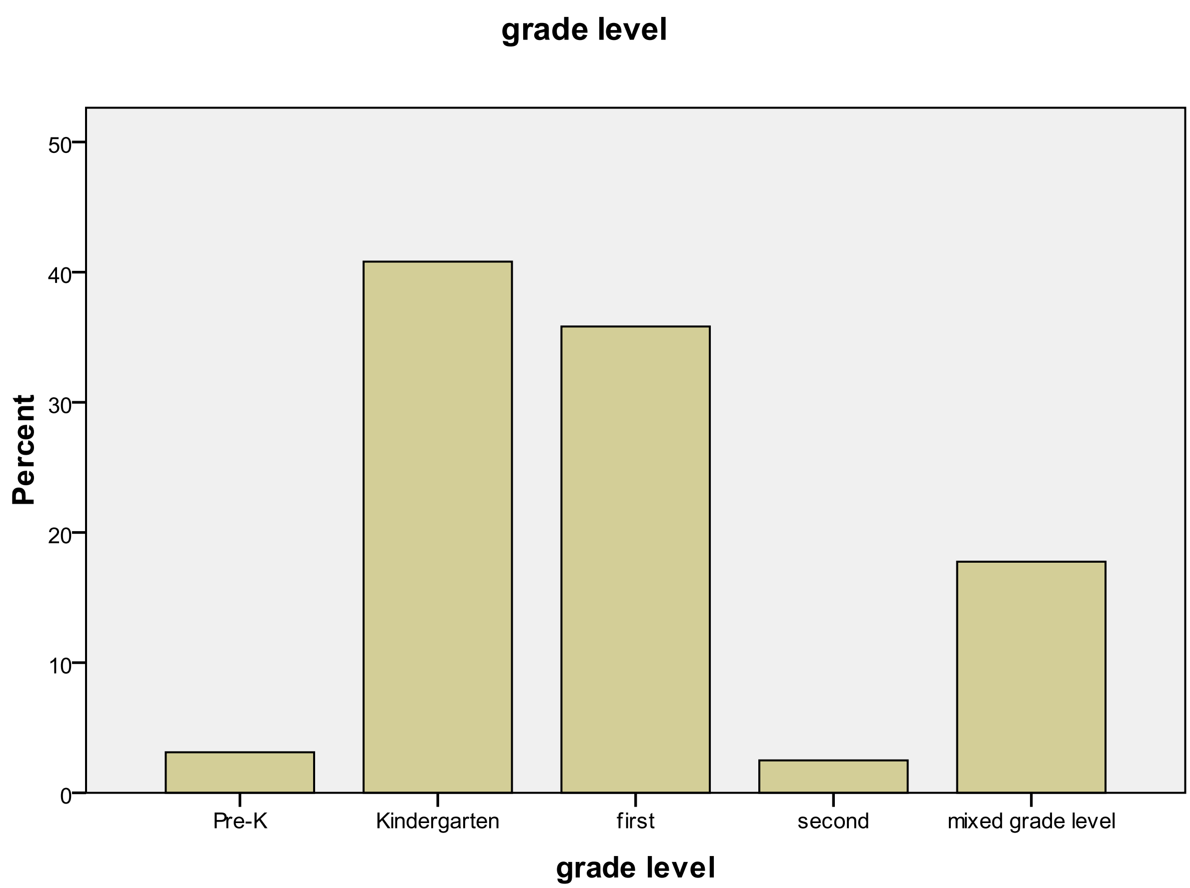
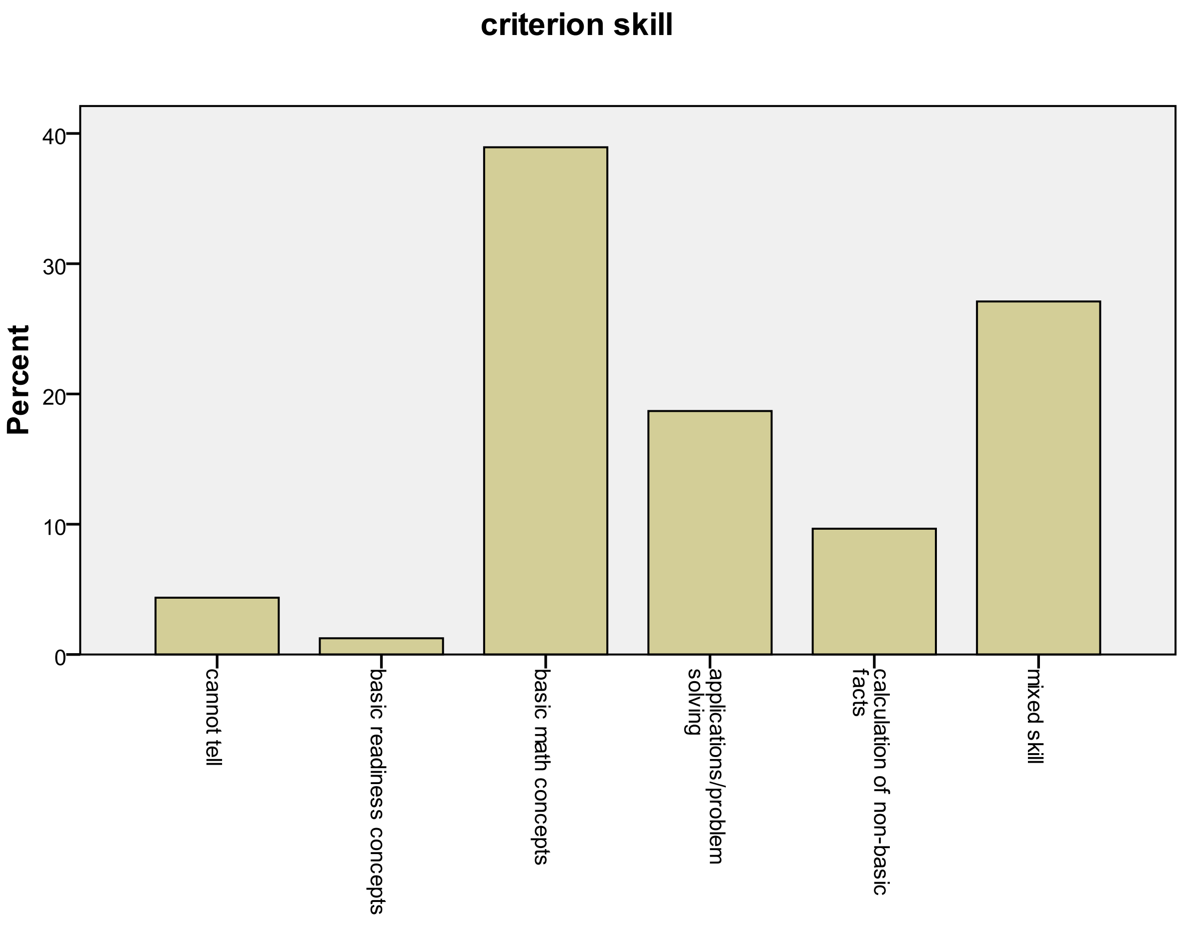

grade level
Percent
50
40
30
20
10
0
Pre-K
Kindergarten
first
second
mixed grade level
grade level

Figure 2.


skill of predictor
40
30
20
10
0
Percent
counting
comparing and ordering
adding to and taking aw ay
mixed or other
readiness concepts
skill of predictor

Figure 3.


criterion skill
Percent
40
30
20
10
0
cannot tell
basic readiness concepts
basic math concepts
applications/problem solving
calculation of non-basic facts
mixed skill

Appendix A: Predictor Measures Guide

Guide to Early Numeracy Measures Abbreviations by Skill Category Included in the Analysis

<u>Counting</u>

C20 – Count to 20

C3 – Count from 3

C6 – Count from 6

CA – Counting Aloud

CB10 – Count by 10s

CB2 – Count by 2s

CB5 – Count by 5s

CO – Counting Objects

COF – Counting on Fluency

CS – Counting Skills

FC – Free Count

MN – Missing Number

NIDC – Number Identification/Counting

NK – Number Knowledge

OC – Oral Counting

OCF – Oral Counting Fluency

OOCCF – One to One Correspondence Fluency

QA – Quantity Array

Comparing and Ordering

OPF – Ordinal Position Fluency

QCF – Quantity Comparison Fluency

QD – Quantity Discrimination

Adding to/Taking Away

FR – Fact Retrieval

NC – Nonverbal Calculation

SP – Story Problems

SSC – Single Skill Computation

Mixed or Other

D – Discrimination

EST – Estimation

MCN – Math Circle Number

MQF – Match Quantity Fluency

MSC – Multiple Skill Computation

MWN – Math Write Number

NI – Number Identification

NID – Number Identification

NN – Number Naming

NNF – Number Naming Fluency

NP – Number Production

NR – Number Reading

NRF – Number Recognition Fluency

NS – Number Sense

NW – Number Writing

RNN – Rapid Naming Numbers

<u>Readiness Concepts</u>

CN – Circle Number

MC – Measurement Concepts

MDC – Math Draw Circle

RNC – Rapid Naming Colors

RND – Rapid Naming Objects

SN – Shape Naming

Appendix B: Criterion Measures Guide

List of Criterion Measures Included in the Analysis

ARITH – Wide-Range Achievement Test 3-Arithmetic

BBCS-Q – Bracken Basic Concepts Scale-Quantitative

BBCS-R – Bracken Basic Concepts Scale-Revised, School Readiness Scale

BRIGANCE – Brigance Screens

CBM-ADD – Curriculum-Based Measurement Addition

CBM-CN – Curriculum-Based Measurement Circle Number

CBM-SUB – Curriculum-Based Measurement Subtraction

CBMCA – Curriculum-Based Measurement Concepts/Applications

CIBS-R – Comprehensive Inventory of Basic Skills, Revised

MBA – Woodcock-McGrew-Werder Mini Battery of Achievement

MCBM – Mathematics Curriculum-Based Measurement

NKT – Number Knowledge Test

SAT-10 – Stanford 10 Achievement Test

SESAT – Stanford Early School Achievement Test

TEMA-3 – Test of Early Mathematics Ability-Third Edition

WJ-III – Woodcock-Johnson, Third Edition

WJ-III-AP – Woodcock-Johnson III Applied Problems Subtest

WJ-III-BS – Woodcock Johnson III math measures Broad Score

WJ-III-CALC – Woodcock-Johnson III Calculation Subtest

WJ-III-FL – Woodcock-Johnson III Fluency Subtest

WJ-III-MATH –Woodcock-Johnson Calculation and Applied Problems Subtests

WJ-III-MR – Woodcock-Johnson III Math Reasoning Subtest

WJ-III-QC – Woodcock-Johnson III Quantitative Concepts Subtest

WJ-R-BM – Woodcock-Johnson Revised- Broad Math

WP –Jordan's Story Problems (Word Problems)

Appendix C: Coding Sheet

Coding Sheet Used to Code the Included Articles

Main					**(EM-CBM) Predictor**												**Criterion Measure**												**Demographic**									**Procedure**			
					test info				desc. data		format		reliability data				test info																								
id	rn	r	rtyp	tmy	pab	Pskl	pcat	ptm	pm	ps	pafl	paf2	ptrr	pintr	pirr	pafr	cab	cskl	cm	cs	cscr	cnrm	caf	ctrr	cintr	cirr	cafr	ctm	n	glv	eth	gnd	nml	nfm	ses	edst	loc	smp	ext	exq	pin

Main Correlation Codes	Predictor Information Codes	Criterion Measure Information Codes	Demographic and Procedural Variable Codes
id: corresponds to study rn: Assign each predictor-criterion correlation (not reliability correlations) a unique number beginning with 1. The correlations can be found in either tables or results sections. r: Report the actual identified concurrent / criterion related correlation coefficient. rtyp: 1 = concurrent (predictor measure administered at the same time or within 2-3 weeks of criterion) 2 = predictive (predictor given one month or more prior to criterion) 0 = cannot tell tmy: 1=fall, 2=wint, 3=spring	pab: write in the abbreviation of the test. pcat: assign 1 if informal and 2 if formal ptm: 1=one minute, 2= more than one but less than 4, 3 = 4 or more. It is important to focus on what the score metric represents. In many cases it is a one-minute test or a units per minute metric. pskl: 1=counting, 2=comparing and ordering,3= equal partitioning, 4=composing and decomposing, 5=grouping and place value,6= adding to and taking away, 7=mixed or other, 8=shapes or other readiness concepts (color, size, pattern, etc). pm: report the mean for the measure ps: report the standard deviation for the measure paf1: 1=individual administration, 2=group administration paf2: 1=latency timing; 2=continuous timing ptrr: report the test-retest reliability coefficient for the measure pintr: report the internal consistency and/or alpha coefficient for the measure pirr: report the inter-rater reliability coefficient for the measure pafr: report the alternate form reliability data for the measure	cab: write in / create the abbreviation of the test cskil: 0=cannot tell, 1=basic readiness concepts, 2=basic math concepts, 3=applications/problem solving, 4=calculation of non-basic facts, 5=mixed skill cm: report the mean cs: report the standard deviation cscr: 0=cannot tell, 1=single sub-test/part of an achievement battery, 2=composite score, 3=single skill test cnrm: 1=nationally normed, 2=state-specific caf: 0=cannot tell, 1=administered in an individual setting, 2=group setting ctrr: report the test-retest reliability coefficient for the measure cintr: report the internal consistency and/or alpha coefficient for the measure cirr: report the inter-rater reliability coefficient for the measure cafr: report the alternate form reliability data for the measure ctm: 0=cannot tell, 1=fall, 2=winter, 3=spring	n: report sample size glv: 1=prek, 2=k, 3=1st, 4=2nd, 5=mixed grade level eth: 0=cant tell, 1=>50%white, 2=black, 3=Hispanic, 4=asian, 5=other, 6=mixed gnd=0=cannot tell, 1=>50%male, 2=female nml: number of males nfm: number of females ses: 0=cannot tell, 1=free & reduced, 2= ineligible for free/reduced lunch edst: 0=cannot tell, 1=>50%gen. ed, 2=sp. ed, 3= ESL loc: 0=cannot tell, 1=Mideast, 2=Midwest, 3=northeast, 4=northwest, 5=southeast, 6=southwest, 7=mixed smp: sample type (0=cannot tell, 1= random, 2=stratified, 3=convenient) ext: 0=cannot tell, 1=examiners were trained, 2=examiners not trained exq: examiner qualification (0=cannot tell,1= pre-professional such as graduate level student, 2=professional such as a teacher, counselor, slp, 3=other, or 4=mixed) pin: procedural integrity(0=cannot tell, 1=used data, 2=no data, 3=narrative)