2016

Quality of Math Instruction Matters: Examining Validity and Reliability of High Impact Strategies in Early Mathematics (his-Em)

Bilge Cerezci
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

QUALITY OF MATH INSTRUCTION MATTERS: EXAMINING VALIDITY AND RELIABILITY OF HIGH IMPACT STRATEGIES IN EARLY MATHEMATICS (HIS-EM)

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN APPLIED CHILD DEVELOPMENT

BY

BILGE CEREZCI

CHICAGO, IL

DECEMBER 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of this study is to introduce a measure of early mathematics teaching practices, the High Impact Strategies in Early Mathematics (HIS-EM) and its framework, to examine its criterion-related, predictive validity of High Impact Strategies in Early Mathematics (HIS-EM) and to describe the types of early math teaching that the HIS-EM detects among a sample of Pre-kindergarten to 3rd grade teachers working with high need students in an urban public schools system. The findings indicate that the HIS-EM produced reliable scores and that meaningful and predictable associations were found between HIS-EM and CLASS. The results also suggested that high quality mathematics as measured by HIS-EM significantly predicted students' mathematics learning at the classroom level. Furthermore, despite the existence of learning standards and increased curricular attention to mathematics, results also revealed that the majority of early childhood educators tend not to provide high quality of mathematics instruction during the course of mathematics teaching. Overall this study shows that the HIS-EM holds promise as a useful tool in mathematics education research, measuring indicators of quality of early mathematics teaching practices.

CHAPTER 1

INTRODUCTION

Mathematics is a means of communication that sharpens people's thinking and their understanding of each other and the world around them. As people advance in their mathematics knowledge and skills, they demand corresponding improvements in the products they use, services they receive, their standard of living, and their country's economy. Lack of mathematics skills, on the other hand, can be an overwhelming obstacle to achieving individual success and improving economic functions.

How can nations make sure their citizens are equipped with the necessary math skills and knowledge in order to be competent and productive members of society? The answer sounds simple, yet is very complex: as early as possible. Growing evidence demonstrates that early mathematics teaching and learning experiences, among all educational resources, are especially important contributors to students' learning and later achievement in mathematics and other areas, particularly in low-SES students who are at risk of falling behind in mathematics achievement (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Carr & Peters & Young-Loveridge, 1994; Lee & Ginsburg, 2009; NCMST, 2000; Sanders & Rivers, 1996; Starkey, Klein, & DeFlorio, 2014).

As early mathematics education has assumed heightened importance, quality of early mathematics teaching and learning experiences has attracted national attention, and the pressure to perform in mathematics has trickled down to preschoolers and kindergarteners. But, how can we provide our youngsters with the necessary skills and knowledge to succeed in math? The National Commission on Mathematics and Science Teaching for the 21st Century raises the same concern by asking a similar question: "As our children move toward the day when their decisions will be the ones shaping a new America, will they be equipped with the mathematical and scientific tools needed to meet those challenges and capitalize on those opportunities?" (NCMST, 2000, p.7).

Committing to equipping every student with high-quality early math learning experiences is undeniably a tall order and an impossible task if it is not approached as being a nationwide issue. Recently in the U.S., there has been an increased emphasis on establishing new standards and principles that could improve quality of mathematics teaching and learning for all students. The National Council of Teachers of Mathematics (NCTM) has published two documents, *The Principles and Standards for School Mathematics* (NCTM, 2000); and *Mathematics Teaching Today: Improving Practice, Improving Student Learning* (NCTM, 2007), which both urge making high quality mathematics teaching and learning a shared experience for all students. These standards and principles suggest that providing mathematics instruction as early as possible may be particularly beneficial if the teachers guide children's mathematical thinking and learning through intentional and explicit teaching. Research also shows that even though some progress has been made at the elementary and middle school levels (National Mathematics Advisory Panel, 2008; Fuson, Carroll, & Drueck, 2000), there is still an unfortunate disparity between the field's vision of quality mathematics teaching and actual mathematics teaching occurring in most early childhood educational settings, which makes NCTM's guidelines all the more timely (Kazemi, Kranke, & Lampert, 2009). Research in cognitive developmental and educational psychology and in mathematics education sheds some light on the underlying reasons behind this disparity by indicating how early childhood teachers often underestimate young children's math knowledge and skills (Clements & Sarama, 2009), prefer to teach other content areas, such as literacy, instead of mathematics (Ginsburg, Lee, & Boyd, 2008; Hausken & Rathbun, 2004), and do not have the content knowledge and skills in mathematics needed to achieve the goals for learning and instruction set out by the NCTM and many other reform programs (Ball 1990; Ma, 1999; Hill, Schilling, & Ball, 2004).

In order to remedy the effects of teachers' misconceptions and beliefs in teaching mathematics to young children, the field needs to go beyond establishing principles and guidelines for teaching mathematics and investigate ways to fully understand what is really happening in early childhood classrooms in the course of mathematics instruction. Existing studies reveal that assessments of early mathematics instruction are needed to identify and improve the quality of early mathematics teaching and education (Boston, 2012). As a response to this need to better understand what quality of instruction entails, several researchers have designed and used a number of instruments to measure and portray the mathematics teaching quality in early childhood classrooms (i.e., surveys methods and conducting observations). While surveys are used efficiently to measure teachers' beliefs and knowledge in mathematics, they are found to be less reliable and consistent when it comes to measuring mathematics teaching practice, due to their subjective nature (Ball & Rowan, 2004; Chen, McCray, Adams, & Leow, 2013; Rowan, Correnti, & Miller, 2002; Spaillane & Zeuli, 1999). In contrast, observing how teachers instruct and provide opportunities for students to learn mathematics revealed more consistent results in understanding teaching quality and its effects on students' mathematics achievement (Boston, 2012; Smith, Lee, & Newmann, 2001; Stigler & Hiebert, 2004). Despite the progress made in quantifying mathematics teaching quality in primary grades, only a handful of researchers have attempted to develop tools to observe early mathematics teaching practices in early childhood settings, specifically in preschool and kindergarten (Learning Mathematics for Teaching, 2014; Piburn Sawada, Falconer, Turley, Benford, & Bloom, 2000; Sarama & Clements, 2007). Even though most of these available observation tools could potentially provide useful information concerning facets of mathematics instruction, each exhibits a varying degree of strength and weakness. More specifically, some tools are not based on a clear conceptual framework and have demonstrated little or no relationship to student learning outcomes, while the reliability and validity of others have not been investigated or reported. As the need increases for better understanding of the quality of teaching practices in mathematics, valid and reliable measures with a strong conceptual framework remain scarce. Thus, there is a need in the field for a reliable observational measure with a strong theoretical framework which would focus on identifying the instructional interactions in mathematics that effectively support young children's early mathematics skills development. Furthermore, tools of this nature can also be used to guide and improve the quality of early math instruction. Such valid and reliable measures can add to the existing body of research on mathematics teaching and offer insights for teachers, teacher educators, and researchers regarding instructional practices in mathematics.

The Current Study

The present study represents an effort to engineer a conceptually-founded, reliable and valid observational tool to measure the quality of early mathematics teaching and examine what happens in classrooms as teachers deliver foundational mathematics content to students. More specifically, the primary goal of this study is to introduce the High Impact Strategies in Early Mathematics, (HIS-EM), an observation tool designed to identify and measure the quality of mathematics teaching practices, and discuss the evidence of its reliability and validity in early childhood classrooms.

Research Questions

The current study is guided by the following research questions:

1) To what extent will constructs measured by HIS-EM and CLASS converge with or discriminate from one another?
2) Does the quality of mathematics teaching as measured by HIS-EM predict children's mathematical gains?
3) What is the profile of early childhood teachers' mathematics teaching quality measured by HIS-EM?

Exploring the quality of mathematics teaching in early childhood classrooms can identify gaps both in mathematics teaching and in how mathematics concepts are taught to young children. Additionally, an observation instrument that is math lesson focused and specifically developed for use in early childhood classrooms can provide more rigorous conclusions of relationships between teachers' math teaching quality and students' math learning outcomes. It may also provide more specific guidance on how to promote foundational mathematics learning. This study ultimately aims to provide meaningful, applicable, and usable information to researchers and teachers, and inform quality of early mathematics teaching and its effects on students' mathematics learning. The proceeding chapter provides an extensive literature review on topics related to early mathematics and teaching. Chapter 3 provides details of study methods and procedures. Study results are summarized in Chapter 4 followed by discussion and interpretation of study results in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

After providing a brief overview of the importance of early mathematics education, subsequent sections are organized around five primary literature reviews. The first literature review focuses on the importance of early math experiences and how teachers can promote foundational mathematics learning through intentional and high quality mathematics teaching. The second literature review summarizes the current issues observed in early mathematics teaching and outlines select studies cited by the review in support of common misconceptions regarding early mathematics teaching and existing lack of knowledge and confidence among early childhood teachers for teaching mathematics. The third literature review focuses on measuring early mathematics teaching quality and summarizing most commonly used methods to study and measure mathematics teaching and its quality. The fourth literature review reviews the observation instruments used in the field to measure early mathematics teaching quality and identifies their strengths and weaknesses, and also highlights the need for conceptually founded, reliable and valid observation to be used in early childhood classrooms to better understand mathematics teaching quality. The fifth and last section introduces High Impact Strategies in Early Mathematics (HIS-EM), is a lesson-based observation tool that is designed to be used in preschool through third-grade classrooms in order to measure the quality of mathematics teaching, specifically its conceptual model and tool characteristics.

The Importance of Early Mathematics Education

Mathematics is a shared universal language and an integral part of everyday experiences for all human beings. It provides insight into the power of the human mind and constitutes the core of any productive economy. The National Commission on Mathematics and Science Teaching (NCMST) postulated there are at least four major reasons that underline the importance of acquiring mathematics competence: (1) the constantly changing demands of the interdependent global economy require and value extensive knowledge of mathematics; (2) American citizens need to have mathematics skills and knowledge in order to compete in that changing economy; (3) knowledge in mathematics is closely related to the nation’s security and future; and (4) “the deeper, intrinsic value of mathematical and scientific knowledge shapes and defines our common life, history, and culture” (2000, p.7). The message from NCMST is clear: mathematics provides the foundation both for lifelong learning and our civilization’s future progress.

The National Council of Teachers of Mathematics (NCTM) advances the claim of mathematics’ importance one step further by connecting it to individual growth and social success and states, “in this changing world, those who understand and can do mathematics will have significantly enhanced opportunities and options for shaping their futures. A lack of mathematical competence keeps the doors closed” (NCTM, 2000, p. 5). In order for our society to develop citizens who are knowledgeable and globally competitive, it is essential to provide them with excellent quality mathematical experiences and facilitate their mathematical abilities.

Unfortunately, schools in the U.S. are not adequately preparing students to meet 21st century demands (Romberg & Kaput, 1999). In fact, “Since the 1970s, a series of assessments of U.S. students’ performances has revealed an overall level of mathematical proficiency well below what is desired and needed” (National Association for the Education of Young Children and National Council of Teachers of Mathematics [NAEYC & NCTM], 2002, p. 1). For example, in the most recent Program for International Student Assessment (PISA) administered by the Organization for Economic Cooperation and Development (OECD), the average math score of American 15-year-olds was 481, 132 points lower than the first country (China), and 13 points lower than the OECD’s average, positioning Americans 36th out of 65 participating countries (OECD, 2012).

American children not only lag behind their peers mathematically, but they also perform poorly on their own national mathematics tests. The 2013 National Assessment of Educational Progress Report (NAEP) suggested that 58% of a nationally representative sample of American students in 4th grade scored below a proficient level in mathematics achievement, and of those, 17% scored below basic level (See Figure 1).

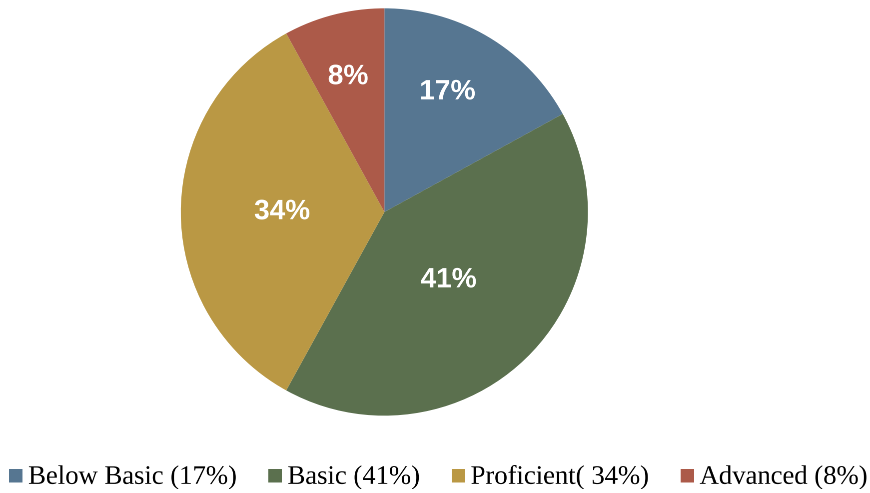

Figure 1. National 4th Grade Mathematics Achievement Reported by NAEP (2013).

This poor mathematics performance demonstrated by American students commences from the time of school entrance. Mounting evidence indicates the dependence of later school performance on the quality of early math experience (Aunola, Leskinen, Lerkkanen, & Nurmi 2004; Carr & Peters, 1995; Duncan, Cleassens, Huston, Pagani, Engel, Sexton et. al. 2007).

If a student falls behind mathematically during the critical years of early schooling, it becomes increasingly unlikely that the student will catch up as she or he moves up the grade levels (Aunola, et. al., 2004; Bodovski & Farkas, 2007). Such research results are both alarming and indicative: early mathematics education is foundational and attention to high quality early math education is vital to improving American students' performance in mathematics.

The Early Math Experience Matters

Early mathematics education refers to teaching that is designed to help young children learn math during their preschool and kindergarten years (Ginsburg, Lee, & Boyd, 2008). Two major research developments have led to growing appreciation of the importance of early math education. First, recent research has shown that young children are able to understand more complex mathematical concepts than was previously thought (Clements & Sarama, 2007; Mix, 2001; Wynn, 1992). Second, research has suggested that early mathematics performance significantly impacts overall school achievement in later life (Aunola, et. al., 2004; Aunio & Niemivitra, 2010) and other subject areas (Denton & West, 2002; Duncan et. al., 2007; Lerkkanen, Rasku-Puttonen, Aunola & Nurmi, 2005). Both developments, described in more detail below, have led to a "call to action" from researchers and policy makers urging that more attention be paid to early mathematics education (National Research Council [NRC], 2009).

Traditionally, mathematics education has not been considered developmentally appropriate for young children (Battista, 1999). Math is abstract while young children are deemed to be concrete thinkers, and some cognitive developmental work done in the mid-twentieth century has been used to suggest that young children's mathematical ideas develop on their own timetable, independent of environmental factors like teaching (Piaget, 1969). Over the past two decades, however, a growing body of literature has indicated that many mathematical competencies, such as sensitivity to set, size, pattern, and quantity are present very early in life (NRC, 2009). Young children have more mathematical knowledge than was previously believed, such as an understanding of number and spatial sense. For example, research suggests that young children have a basic understanding of one-to-one correspondence even before they can count verbally (e.g., pointing to items in a collection and labeling each with a number) (Mix, 2001). Furthermore, young children also enjoy exploring spatial positions and attributes of geometric shapes by building towers with blocks and cubes and by manipulating various materials, such as puzzles and two and three dimensional shapes (Clements, 1999; Clements & Sarama, 2008). They also demonstrate emerging awareness of measurement, beginning to notice and verbalize similarities and differences in the size, height, weight and length of various objects and materials (Clements & Sarama, 2008). In addition, research also suggests that three and four year-old children engage in analytical thinking as they collect and sort materials by various attributes (e.g., color, size, and shape) and employ algebraic thinking as they copy patterns observed in their surroundings and create their own by using pattern blocks and other materials (Epstein, 2003; 2006). In fact, as research states, most children enter school with a natural wealth of knowledge in early mathematics and cognitive skills that provide a strong foundation for mathematical learning (Clements & Sarama, 2009; Ginsburg, Lee, & Boyd, 2008; Mix, 2001).

New evidence also indicates that achievement in early mathematics has a profound impact on later success. A longitudinal study by Aunio & Niemivitra (2010) with 212 Finnish kindergarten children examined whether children's mathematics skills in kindergarten can predict their mathematics performance in the first grade. The results suggested that specific mathematics skills such as counting in kindergarten are associated with learning basic and applied arithmetic skills and the overall quality of mathematics achievement in the first grade. Another study done by Aunola and his colleagues investigated how children's math development occurs from Pre-K to Grade 2. Over the course of three years, the researchers worked with 194 Finnish children whose math performances were examined twice each year. The results suggested that differences among children's math performance increase over time and these discrepancies exist as early as preschool years (Aunola, et. al., 2004). Based on the results, the authors claimed, "Children who entered preschool with a high level of math skills showed rapid development later on, whereas those who started at a lower skill level showed relatively slower development" (Aunola, et. al., 2004, p. 711).

The impact of early math skills is not limited to academic achievement in primary grades but carries on through high school and beyond (Duncan & Magnuson, 2009; 2011; Entwisle & Alexander, 1990; NRC, 2009; Stevenson & Newman, 1986). For example, Duncan and Magnuson (2009) examined the mathematics achievement of children who consistently exhibited persistent problems in understanding mathematics in elementary school and analyzed it in comparison to children who had stronger early math abilities. The results of the study revealed that 13% of the children with persistent problems were less likely to graduate from high school and 29% of them are less likely to attend college than those who had stronger early mathematics abilities (see Table 1). In other words, the initial differences in mathematics skills in early years may lead children to lag behind their more knowledgeable peers not only in primary grades but throughout their formal schooling (Geary, Hoard, & Hamson, 1999).

Table 1. Effect of Persistent vs. No Problems in Mathematics Compared to Other Areas at Ages 6, 8 and 10 on the Probabilities of High School Graduation and College Attendance (Adapted and redrawn with permission, Duncan & Magnuson, 2009).

Problem Area	High School Completion	College Attendance
Reading	-.05	-.06
Mathematics	-.13*	-.29**
Anti-social Behavior	-.10	-.24*
Inattention	.01	-.05
Anxiety	-.03	-.18

Note: ** p<.01 *p<.05

Studies also showed the predictive power of early math skills compared to other academic skills, such as reading. Lerkkanen, Rasku-Puttonen, Aunola and Nurmi (2005) investigated the relationship between mathematical performance and reading comprehension among 114 seven-year-old Finnish-speaking children during the first and second years of primary school. The results stressed that the importance of the mathematical knowledge children have before schooling is very important because these skills are predictive of their subsequent reading comprehension. In other words, early mathematics skills predict not only later achievement in mathematics but also later reading achievement.

Similarly, Duncan and colleagues (2007) conducted a meta-analysis of six large-scale longitudinal data sets to examine the relationship between early learning and later school achievement. Of these, two were nationally representative of U.S. children, two were gathered from multi-site studies of U.S. children, and the last two focused on children either from Great Britain or Canada. The researchers focused on the relationship between school-entry skills (i.e., reading achievement, math achievement, attention, internalizing behavior problems, social skills, and anti-social behavior) and later math and reading achievement while controlling for children's preschool cognitive ability, behavior, and other important background characteristics such as socioeconomic status, mother's education, family structure, and child health. Their meta-analysis revealed that only three of the six sets of school entry skills and behavior are predictive of school achievement: math, reading, and level of attention. Further, early math skills were consistently a stronger predictor of later achievement compared to reading and level of attention (Duncan, et. al., 2007). Consistent with the educational attainment analyses (Duncan & Magnuson, 2009), early math achievement was found to be the most powerful predictor of later school achievement (Duncan, et. al., 2007).

The above two areas of recent research are important contributions to the field of early mathematics education. The first expands our knowledge of young children's capacity to learn mathematics and challenges early childhood educators to find ways to support and nurture such capacity in developmentally appropriate ways (Clements & Sarama, 2009; Ginsburg, Lee, & Boyd, 2008; NRC, 2009). The second area of research illustrates the importance of early mathematics education as it can indeed provide children with a distinct educational advantage in later years (Duncan et al., 2007; Griffin, Case, & Siegler, 1994; NRC, 2009). These research findings urged researchers, educators, and policy makers to respond with a series of calls to action: it is of critical importance that early childhood educators pay greater attention to early math education and ensure that quality early math teaching takes place in the classroom (Barnett, 2008; Clements, Sarama, & DiBiase, 2004; NRC, 2009).

Early Math Teaching Matters

Even though young children are natural mathematicians (NRC 2009) and capable of developing some complex mathematical ideas (e.g., addition) and strategies (e.g., sorting by multiple attributes to analyze data), it is also true that they do not become skilled in mathematics without intentional and high quality instruction (Baroody, 2001; Baroody & Dowker, 2003; Clements, 2001; Epstein, 2003; Richardson & Salkeld, 1995).

Intentional teaching in early mathematics education refers to teaching that is carried out with specific mathematics learning outcomes or goals in mind to support children's understanding and learning of mathematics. Intentional early math teaching matters because it assists young children to bridge the spontaneous, or informal knowledge that they acquire through daily experience with the scientific, or formal knowledge that will serve them for the rest of their academic studies and beyond (Vygotsky, 1978). Young children from birth to age five have developed a range of informal mathematics knowledge, including ideas of more or less, shapes, patterns, measurement, the meaning of numbers, and how numbers work (Clements, 2001; Gelman, 2000). For example, a child as young as two knows if she gets more or less crackers than her friend next to her. She also knows that her dog is bigger than her cat, and that nursery school is much closer than the grandmother's house. While serving as important building blocks, such informal and intuitive mathematics understanding does not necessarily help young children explicitly examine and interpret their experiences in mathematical forms. Mathematics' representational systems, formulas, theorems, and procedures, are artifacts of man-made knowledge, the result of thousands of years of human exploration, experimentation, and innovation (NCMST, 2000). Of critical importance to the acquisition of such man-made knowledge is intentional instruction. Lee and Ginsburg summarized this point well, “Children do indeed learn some mathematics on their own from free play. However, it does not afford the extensive and explicit examination of mathematical ideas that can be provided only with adult guidance.” (2009, p. 40).

Intentional mathematics teaching means more than arranging the classroom materials, utilizing random teachable moments, or just counting the days of the week as a mathematics activity. To be mathematically intentional, “teachers must set aside time specifically for the study of mathematics and be purposeful in planning experiences that help children develop specific mathematical understandings” (Richardson & Salkeld, 1995, p. 42). Intentional teachers must also be professionals who keep in mind the key goals for children's learning and development in early mathematics. They do not only ensure that teaching builds on the mathematical ideas and skills that the children already possess, but also allow their students to extend their thinking further by creating supportive environments, planning curriculum, and selecting from a variety of teaching strategies that best promote each child's thinking and skills (Clements, 2001; Perry & Dockett, 2002).

The provision of intentional instruction in early mathematics, though necessary, is not sufficient by itself to facilitate mathematical proficiency. Instructional interactions must also be of high quality. Quality mathematics teaching refers to mathematics instruction that meets the demands of the mathematics discipline as well as enables all students to engage in meaningful, conceptual and developmentally appropriate mathematics learning through effective instructional support in mathematics. Intentional teachers would be more likely to provide high quality teaching and learning experiences to young children by purposefully designing learning opportunities that encourage children to explicitly think, talk and act on real-life experiences and problems in mathematical ways (Doabler, Baker, Kosty, Smolkowski, Clarke, Miller, & Fien, in press). When intentional teaching includes quality teaching, children will have a chance to meaningfully engage in foundational mathematical principles and develop robust and transferable knowledge in mathematics (Ginsburg, 2009).

For example, in an effort to assess the quality of mathematics teaching in U.S. classrooms and document how variations in quality of teaching might produce different student outcomes, Rivkin and colleagues (2005) collected and analyzed the math test scores of approximately one-half million students in grades 3 through 7 at over 3000 schools in Texas (Rivkin, Hanushek, & Kain, 2005). In addition, specific demographic characteristics of teachers and schools (e.g., teacher experience, education, and class size) were also collected in order to estimate the quality of teaching. The researchers used matched data on teachers and students and made estimations in variations in teaching quality. Quality of instruction provided by the studied teachers was categorized as “low” or “high.” The final report on this study suggested that students whose teachers provided high quality instruction gained 1.5 grade equivalents while students whose teachers provided low quality instruction only made a gain of 0.5 grade equivalents during the same academic year. Weiss and Pasley (2004) also highlight the importance of high quality teaching in building foundations for future learning by stating that high quality teaching is more likely to engage students with important mathematics content and build on students' capacity to understand and implement these mathematics concepts.

While Rivkin and his colleagues reported on the effects of quality of instruction on student achievement in a single school year, Sanders and Rivers (1996) emphasized the cumulative effects of quality of mathematics teaching. In their study, Sanders and Rivers (1996) developed a value-added model to measure individual teacher contributions to student learning and investigated the long-term effects of teachers on the mathematics achievement of 5th grade students in two Tennessee districts. By grouping teachers into quintiles according to the size of their previous students' achievement gains, the researchers estimated how being assigned to different teachers with varying levels of effectiveness would influence the students' achievement in mathematics. The results of this study suggested that both ineffective and effective teachers have additive and cumulative effects on student achievements in mathematics. Further, these effects were not compensatory. That is, the disparities in student performance as the result of varying teaching qualities in mathematics show a persistent pattern or have an enduring effect either for better or worse (See Figure 2).

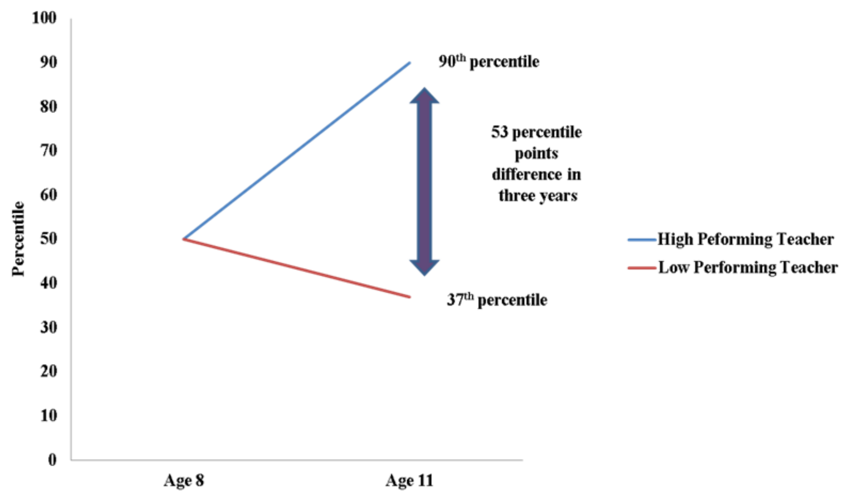

Figure 2. Student Performance over Time as the Result of Varying Teaching Quality

Furthermore, intentional early math teaching matters because it helps narrow the achievement gap. It is well documented that the mathematical skills of young children from low-income families lag behind those of their middle-income peers (Flanagan, McPhee, & Mulligan, 2009; Starkey, Klein, & Wakeley, 2004). For example, in a recent assessment of school readiness, by the Chicago Department of Children and Family Services, among 7,354 kindergarten-age children attending Head Start programs in Chicago, 2,059 of them, or nearly one third, performed below the standard for school readiness in kindergarten mathematics based on the Teaching Standards GOLD Assessment System (Chicago DFSS, 2014). The gap between low and middle-income children includes a wide range of early mathematical areas, such as number sense, spatial sense and geometry, and measurement (Clements, Sarama, & Gerber, 2005; Klein & Starkey, 2004; Saxe, Guberman, & Gearheart, 1987). Intentional and high quality early math teaching, however, can indeed narrow the achievement gap, as exemplified by the work of Starkey and colleagues (2014) who conducted a cluster randomization study to measure effectiveness of "Pre-kindergarten Mathematics" intervention on low SES children's early mathematical development (Starkey, Klein, & DeFlorio, 2014). The goal of the intervention was to help teachers engage in intentionally planned mathematics teaching activities to support the development of children's informal mathematical knowledge. In this study, a total of 63 Head Start classrooms at 43 sites serving an urban, ethnically diverse, low-income population, were randomly assigned to one of three conditions: 2-year intervention group, 1-year intervention group, or a control group (Starkey et. al., 2014). The three conditions differed in terms of the number of years of math intervention that the children and the teachers received. While the 2-year intervention group received an intervention in Pre-Pre-K Mathematics during the first year of preschool and Pre-K Mathematics during the second year, the 1-year intervention group only received Pre-K Mathematics intervention during the second year of preschool. The control group did not receive any Pre-Pre-K Mathematics or Pre-K Mathematics intervention at any point of pre-schooling.

The results of the study suggested intervention group children made significant gains in mathematics compared to children who did not receive any intervention (Starkey, et. al., 2014). In other words, by providing intentional and quality early mathematics instruction, preschool and kindergarten teachers can help disadvantaged children build a strong foundational mathematics knowledge compared to their counterparts who do not receive such instruction and narrow the mathematics achievement gap that young children might face during formal and after schooling (Bahr

& de Garcia, 2010; Rouse, Brooks-Gunn, & McLanahan, 2005).

In summation, recent statistics reported by the U.S. Department of Education (USDOE) and National Center for Education Statistics (NCES) suggest that early childhood education is increasingly becoming a common experience for young children in the U.S. According to the report, between 1980 and 2013 enrollment in some type of school (including private childcare centers, publicly supported preschools and kindergartens, and Head Start) increased substantially: enrollment grew from 27 to 39 percent for three year olds, from 46 to 65 percent for four year olds and from 85 to 88 percent for five year olds (Aud, Wilkinson-Flicker, Kristapovich, Rathbun, Wang, & Zhang, 2013). Given the facts that low mathematics achievement among the U.S. students has its roots in early years, and there are an increasing number of students enrolling in early care and the educational system, the role of teachers in early childhood settings becomes even more essential in terms of providing a robust mathematics experience through intentional and high quality mathematics teaching. Even though young children engage in a substantial amount of mathematical activities and discoveries on a daily basis, evidence indicates that young children, especially those who are more disadvantaged compared to their peers, need intentional mathematics instruction to connect their intuitive mathematics discoveries and knowledge into generalizable and more sophisticated mathematics knowledge (Bahr & de Garcia, 2010; Starkey, Klein, & DeFlorio, 2014; Rouse, Brooks-Gunn, & McLanahan, 2005). Through intentional, high quality, challenging, and accessible mathematics education, preschool and kindergarten children can develop robust mathematics knowledge which will potentially have long-term positive implications for their academic performance in mathematics and other subject areas (Sanders & Rivers, 1996; Weiss & Pasley, 2004). Unfortunately, there is a wide gap between the current reality of early mathematics teaching and desired goals (Ginsburg et al., 2008).

Current Issues in Early Mathematics Teaching

Mounting evidence provides a compelling argument that intentional and high quality mathematics teaching in early childhood years matters and early childhood teachers are in a unique position to reduce the achievement gap in early mathematics and to help all children build a strong foundation in early mathematics. The National Council of Teachers of Mathematics (NCTM) drew on this large body of literature and outlined a broad vision for mathematics teaching by suggesting standards: the mathematical content (Number and Operations, Algebra, Geometry, Measurement, and Data Analysis and Probability) and; the processes (Problem Solving, Reasoning and Proof, Communication, Connections and Representation) in which students should engage and learn. Furthermore, it also suggests six principles to describe features of high-quality mathematics education; (1) high expectations and strong support for all students (Equity); (2) a coherent curriculum of important mathematics, articulated across grade levels (Curriculum); (3) teachers who understand what students need to learn and challenge and support them (Teaching); (4) instruction that connects prior knowledge with new knowledge (Learning); (5) meaningful and intentional assessment that is useful for both teachers and students (Assessment); and (6) use of technology that enhances students' mathematics learning (Technology) (NCTM, 2000).

The NCTM is clear: to ensure intentional and high quality mathematics teaching and learning becomes the norm for all, teachers are not only expected to have a deep conceptual understanding of what early mathematics entails, but also be familiar with developmental trajectories for mathematics learning. Furthermore, teachers are also expected to be intentional in using various tools to teach a range of developmentally appropriate math content, and adjust their teaching based on their students' needs and developmental abilities. Last but not least, teachers are expected to conduct meaningful assessments in order to document their students' progress and use technology as a tool to aid the teaching and learning process.

Unfortunately, there is a great disparity between the teaching principles and standards that are outlined by NCTM and the teaching that is occurring in most U.S. early childhood classrooms (Kazemi, Kranke, & Lampert, 2009). Studies over the past decades consistently reveal two types of issues among the main contributors to this disparity; (1) teachers' misconceptions about early math; and (2) teachers' lack of knowledge and confidence in teaching mathematics.

Common Misconceptions around Early Mathematics Teaching

Myths abound among early childhood teachers regarding early math, including the ideas that early math is not as important as other subject areas such as literacy that young children cannot do mathematics, and that mathematics cannot be taught to young children. These misconceptions often hold teachers back from understanding new principles and standards in early mathematics teaching and may even hinder them from incorporating new vision of early mathematics teaching in their practice.

Young children's math knowledge and ability to learn is underestimated. In recent decades, many researchers have focused on what even young children can do, and have accumulated a wealth of evidence that they exhibit a range of mathematical abilities and competence in mathematics. As Vygotsky (1978) stated, "Children's learning begins long before they enter school ... They have had to deal with operations of division, addition, subtraction, and the determination of size. Consequently, children have their own preschool arithmetic, which only myopic psychologists could ignore" (p. 84). Even though research clearly suggests that young children are capable and motivated to explore mathematical concepts before formal schooling starts (Clements, 1999; Clements & Sarama, 2008; Epstein, 2003; 2006; Mix, 2001), most early childhood teachers underestimate what young children know and what they can learn in mathematics (Brown, 2005; Clements & Sarama, 2009; Graham, Nash, & Paul, 1997; Lee, 2004; Tudge & Doucet, 2004; Van den Heuvel-Panhuizen, 1990).

For example, in one study, Van den Heuvel-Panhuizen (1990) asked groups of preschool teachers and school staff who worked with preschoolers to estimate their preschoolers' mathematical competencies when they entered kindergarten the following year. Results of the study revealed that teachers and staff highly underestimated the math competencies of these young children. Particularly, when more than 80% of these kindergarteners were able to count out nine marbles, the adults' estimates only ranged between 20% and 50%. Further, while more than 40% of these students were able to subtract 8 from 10 without using any manipulatives, adults estimated less than 10% of them would be capable of completing this task (Van den Heuvel-Panhuizen, 1990). This sort of underestimation often compromises what early childhood teachers teach and how they teach it (Brown, 2005; Graham, Nash, & Paul, 1997; Lee, 2004; Tudge & Doucet, 2004). Stated by Lee and Ginsburg, “Teachers often limit their focus to one-to-one correspondence, simple counting and numbers, and perhaps naming and sorting simple shapes, even when children are capable of learning far more complex content” (2009, p.39). While acquiring the basic skills in mathematics is important in early years, teachers need to help children build upon and extend them to deeper and broader mathematical concepts (Clements, 2004; Sarama & Clements, 2010).

Integrated approaches to teaching are always best. The early childhood field has long-favored curriculum integration as a teaching approach in the education of young children, especially in mathematics. Emphasis given to teaching mathematics as a stand-alone subject varies across settings, but is generally minimal in the earliest years (Chung, 1994; Clements & Sarama, 2011; Rudd, Lambert, Satterwhite, & Zaier, 2008). Favoring integrated curriculum over stand-alone subject teaching can be seen in statements such as “because a subject-matter approach to the curriculum is expert-based, much of the content is difficult for children to understand” (Jalongo & Isenberg, 2000, p. 205), and the convictions that “times are set aside to teach each subject without integration” (Bredekamp & Copple, 1997, p. 130) are developmentally inappropriate practice in early childhood education. This vision not only undermines young children’s knowledge and capabilities in early mathematics, but also discourages teachers from teaching mathematics and affects young children’s access to it. As Clements and Sarama noted “Early childhood teachers often believe they are ‘doing mathematics’ when they provide puzzles, blocks, and songs. Even when they teach mathematics, that content is usually not the main focus, but it is embedded in a fine-motor or reading activity” (Clements & Sarama, 2011, p.968). For example, Chung (1994) conducted a study to document the amount of time kindergarten teachers spent on teaching mathematics on a daily basis. Based on the observations gathered from 30 public school kindergarten teachers, Chung (1994) concluded that observed teachers spent about one fourth of their classroom time on teaching mathematics that was usually integrated with other learning activities, and that mathematics was seldom taught as a separate subject. In another study, Rudd and colleagues observed 11 teachers who worked with children ranging in age from birth to five years. Researchers gathered 40 hours of observations in which they noted no incidence that could be identified as intentionally planned mathematics activities (Rudd, Lambert, Satterwhite, & Zaier, 2008).

Even though the integrated approach to teaching mathematics can allow students to investigate the connections between various subjects, it can prevent students from focusing on specific math ideas in a detailed manner, if it is not balanced with a subject-specific mathematics teaching approach. “The curriculum should not become, in the name of integration, a grab bag of any mathematics-related experiences that seem to relate to a theme or project” (NAEYC & NCTM, 2002, p. 8).

Literacy is more important than mathematics. Research also suggests that early childhood teachers tend to emphasize content areas such as language and early literacy at the expense of mathematics education (Early, Barbarin, Bryant, Burhninal, Chang, & Clifford, 2005; Hausken & Rathbun, 2004; Layzer, Goodson, & Moss, 1993).

A study by Early and his colleagues suggested that preschool teachers usually devote more time to teaching literacy activities (21% of classroom time) than mathematics teaching (8% of classroom time) (Early et al., 2005). Similar scant attention given to teaching early mathematics has also been observed among kindergarten teachers. Hausken and Rathbun (2004) found that while kindergarten teachers spend 3.1 hours a week on teaching mathematics, they devote a total of 5.2 hours to reading in general. The amount of time kindergartners spent in mathematics varied by the type of kindergarten program they attended, with kindergartners in full-day programs spending more time in mathematics than their peers in half-day programs; about 3.6 hours per week in full day and about 2.4 hours a week in half-day programs. As a matter of fact, "…mathematics seems to be seriously overlooked in preschool classrooms even when the teachers say that it is important and that they teach it" (Ginsburg, Lee, & Boyd, 2008, p.11).

Lack of Knowledge and Confidence for Teaching Mathematics

Last but not least, in the field of early education, most of the teachers do not possess the mathematical knowledge that is necessary to provide quality mathematics teaching and learning opportunities to young children (Ball, 1990; Ma, 1999; Hill, Schilling, & Ball, 2004) and often do not feel confident in teaching them mathematics (Bursal & Pazkanos, 2006; Copley, 2004; Wilkins, 2008). Teachers cannot teach what they do not know. Further discussing how teachers' content knowledge can affect the quality of their instructional practices, Brophy (1991) states:

> Where (teacher's) knowledge is more explicit, better connected and more integrated, they will tend to teach the subject more dynamically, represent it more varied ways, and encourage and respond fully to student comments and questions.

> Where there knowledge is limited, they will tend to depend on the text content, de-emphasize interactive discourse in favor of seatwork assignments, and in general, portray the subject as a collection of static, factual knowledge. (Brophy, 1991, p.352)

Although the education system highly depends on the work and knowledge of early childhood teachers to help young children learn math concepts and develop math understanding, the same system does not put enough effort into equipping teachers with the necessary mathematics knowledge-base and skills that they require to undertake the task. Research indicates that most of the early childhood programs in higher education do not offer courses specifically devoted to mathematics teaching and learning in early childhood classrooms (Armstrong, Ginet, & Warisi, 2012; Ginsburg, Lee, & Stevenson, 2008; NRC, 2009). Even when they do, it usually does not exceed more than one course, which is not enough to equip prospective teachers with necessary domain specific knowledge in mathematics that they need in order to provide quality mathematics education for preschool and kindergarten children (Copple, 2004; Ginsburg et al., 2008; Ginsburg, Jang, Preston, VanEsselstyn, & Appel, 2004).

To put this argument into perspective, Lobman and colleagues (2005) investigated the contents of the courses offered in early childhood programs in the New Jersey area (Lobman, Ryan, & McLaughlin, 2005). The results suggest that only 16% of preschool to 3rd grade early childhood education programs in New Jersey four year colleges offer coursework specifically allocated to mathematics while 10% do not offer mathematics education at all. Further, 74% offer mathematics education not even as a stand-alone subject course but as a part of a comprehensive early childhood education course. The information about two year community colleges is also disappointing; 18% of them do not offer early childhood mathematics and almost 50% offer in conjunction with another course (Lobman, et. al., 2005).

As McCray and Chen state, “This educational lack is both compounded by and compounds a lack of confidence in their mathematical abilities among early childhood teachers…” (2011, p. 256). Teachers in early childhood education often do not feel confident in their personal knowledge of mathematics and ability to teach it to young children (Bursal & Pazkanos, 2006; Copley, 2004; Wilkins, 2008). Mathematics is often described as their “enemy” or “something I hate” (Cady & Rearden, 2007). In a study conducted by Stipek and colleagues, teacher confidence in mathematics has been found to be highly correlated with students’ learning and students’ confidence in themselves as mathematics learners (Stipek, Givvin, Salmon, & MacGyvers, 2001). Specifically, results suggested that early childhood teachers who enjoy mathematics and feel comfortable teaching mathematics as a subject area tend to encourage their students to engage in problem solving activities more than teachers who feel less confident in teaching mathematics. Low-confidence teachers tend to ignore wrong answers or misconceptions and often give feedback that conveys low expectations for the students (e.g., *I did not expect you to get that right*). Low confidence in teaching mathematics can potentially hinder the teacher’s teaching performance and even influence how confident her students feel as mathematics learners.

To sum up, quality of early mathematics teaching entails a host of professional competencies supported by the NCTM’s Principles and Standards. In this vision, “high quality” mathematics instruction involves intentional teaching, knowing foundational mathematics concepts, being able to help students construct this knowledge, recognizing misconceptions and misunderstandings, providing accurate and supportive feedback, and using a range of tools and representations appropriate to the concepts because they serve as an essential vehicle for teaching children fundamental concepts and skills in early mathematics. Unfortunately, understanding these qualities and portraying them accurately pose a challenge to the field. Current issues observed in early mathematics teaching (e.g., underestimating young children's math skills and knowledge, lack of content knowledge and confidence in teaching mathematics, and etc.) can often interfere with fully understanding and interpreting these principles of early childhood mathematics education (NAEYC & NCTM, 2002), and even hinder preschool and kindergarten teachers from implementing them (Ginsburg et. al., 2008). Finding a way to examine the mathematics instructional quality in early childhood classrooms is a crucial first step to remedying the effects of these problems and promoting high quality mathematics instruction in the early years. Therefore, the ability to measure instructional quality in early mathematics is critically required in efforts to assess, and ultimately, improve instruction.

Measuring Early Mathematics Teaching Quality

Currently, forty-four states in the U.S. offer some form of standards for early mathematics education that emphasize the importance of intentional and quality teaching through which all young children can engage and learn core mathematical concepts (Achieve, 2013; Clements, Sarama and DiBiase, 2004; Cross, Woods, & Scweingruber, 2009; NAEYC, 2010; NCTM, 2000). The question is no longer whether children should be taught math in preschool or kindergarten, but how we can ensure that all children can benefit from the standards set by the National Council of Teachers of Mathematics (NCTM) and other research-based sources for high quality early mathematics teaching and learning experiences.

Examining early mathematics teaching practices in more detail and depth by using measurements of instructional quality at the classroom level, based on the activities that the students and teachers are engaged in during early math lessons, can be a first crucial step in finding an answer to this question. In order to achieve that, however, the field needs to go beyond describing high quality mathematics teaching standards, to incorporating valid and reliable measures that will monitor the quality of early mathematics teaching to make sure that these principles come into life in every early childhood classroom. In the following passage Boston (2012) summarizes this point well

> By capturing what teachers and students are doing in mathematics classrooms in the process of teaching and learning mathematics, measures of instructional quality can identify instructional factors that influence students' learning and uncover important differences in students' opportunities to learn mathematics across classrooms, schools, and districts (p.77).

Such measures can add to the existing body of research on mathematics teaching and offer meaningful implications for teachers, teacher educators, and researchers regarding the quality of instructional practices in early mathematics and its effects on student outcomes, as long as they are sufficiently valid and can be used reliably. Therefore, in order to make sure the early childhood education community meets standards set by the field and that all young children receive the early mathematics education envisioned by the field, appropriate instruments must be used to measure the quality of early mathematics teaching.

Teaching is too complex for any single measure to accurately capture and represent its entirety and quality. Researchers and educators have attempted to design and use a number of instruments to measure and portray the mathematics teaching quality in early childhood classrooms. Existing tools focus primarily on using survey methods and conducting observations.

Using Survey Methods to Study Mathematics Teaching

Survey methods involve collecting information from a sample of individuals through their responses to a set of questions (Johnson & Christensen, 2012). Surveys are popular among many researchers because they are one of the least expensive data collection methods (Paterson, Potoski, & Capitano, 2002). They are quick, relatively easy to administer to a large number of people, and do not depend on the use of sophisticated methodology or equipment (Miller & Hays, 2000).

In recent years, many scholars have used large-scale survey techniques to measure various characteristics of teaching and its relationship with student outcomes in mathematics (Cohen & Hill, 1998; Engel, Claessens, and Finch, 2013; Hill, Rowan & Ball, 2005; Hill, Schilling, & Ball, 2004; Malara & Zan, 2002; Porter, Blank, Smithson, & Osthoff, 2005; Spillane & Zeuli, 1999). While survey methods have been used efficiently and effectively to measure teachers' content knowledge in mathematics (Hill, Rowan & Ball, 2005; Hill, Schilling, & Ball, 2004) and teachers' beliefs in teaching early mathematics to young children (Chen, McCray, Adams, & Leow, 2013; Ma, 1999), mixed evidence exists regarding the extent to which survey-based measures of mathematics teaching practices accurately predict mathematics teaching quality and its effects on student achievement (Ball & Rowan, 2004; Rowan, Correnti, & Miller, 2002; Walkowiak, Berry, Meyer, Rimm-Kaufman & Ottmar, 2013).

Most studies that examined the relationship between teacher-reported classroom practices in mathematics and student achievement have relied on teacher surveys that reported the frequency of specific activities such as cooperative learning groups, use of manipulatives, asking open-ended questions, and so on. These activities are often referred as "standards-based" or "reform-based" practices that are intended to further children's mathematics understanding and development (Cohen & Hill, 2000; Engel et. al., 2013; Hamilton, McCaffrey, Klein, Stecher, Robyn, & Bugliari, 2003). For example, in their study, Cohen and Hill (2000) asked teachers to take a survey about how often they use reform-based mathematics instruction while teaching mathematics (see Table 2). The results of the study revealed that teachers differed in terms of their use of reform-based practices in mathematics teaching and teacher-reported frequency of these practices were positively correlated with mathematics test scores of fourth-graders.

Table 2. A Sample Item from Teacher Survey of Framework Practices. Adapted with permission from Cohen & Hill, 2000 (redrawn)

9. About how often do students in your class take part in the following activities during mathematics instruction? (CIRCLE ONE ON EACH LINE.)	Never	A few times a year	Once or twice a month	Once or twice a week	Almost daily
Make conjectures and explore possible methods to solve a mathematical problem…					
Discuss different ways that they solve particular problems…					
Work in small groups on mathematics problems. . .					
Work on individual projects that take several days. . .					
Work on group investigations that extend for several days. . .					
Write about how to solve a problem in an assignment or test. . .					
Do problems that have more than one correct solution. . .					

In another study, Spillane and Zeuli (1999) studied 25 teachers and used survey methods to investigate the degree of the alignment between their mathematics teaching practices and standards-based practices outlined by the NCTM. Unlike the previous study, the researchers also interviewed teachers to investigate whether they really understood the reform-based mathematics practices. Analyses of the results documented the disparity between what teachers reported and what they actually provided during mathematics teaching. Furthermore, the results also revealed that only four out of twenty-five teachers really understood the core ideas of the standards-based (i.e., NCTM-aligned instructional approaches) mathematics instruction and actually appeared to use them in their classroom while the rest of the teachers appeared to use these terms but did not necessarily apply them in their practice. Spillane and Zeuli (1999) illustrate this point well by stating:

> Ms. Townsend stated that her students "do a lot of discussion" because it gives them the ability to question and explain [their thinking]." What was striking, however, was when Ms. Townsend referred to how discussion was instantiated in her teaching, she gave an example that concerned the rules for a board game

> students were designing: “I was trying to think of an example. They had a [board game] rule that if a student did something, then you lose three turns. And I said, “how does your partner move when they get those extra turns? They hadn't thought of that part yet.” (p.12)

As in the case of Ms. Townsend, many teachers in the study seemed to be good at explaining what they do in the classroom by using terms that imply standards-based practice, but often taught in ways that were not necessarily reform-oriented or standards-based.

Inconsistent results observed in the field can be partially due to the fact that surveys rely heavily on words and phrases to describe instruction even though the language for instruction in mathematics in the U.S. can be very imprecise can vary across teachers, schools, and district. This ultimately makes the use of surveys in measuring teaching practices problematic because survey items can be subject to misinterpretation or misrepresentations of the classroom practices being assessed and might not accurately represent the actual classroom practice or its quality (Spaillane & Zeuli, 1999; Ball & Rowan, 2004; Walkowiak et. al., 2013). For instance, while some scholars have documented variations in teachers’ understanding of terms such as “investigate” or “discuss” and how they use these terms to define various methods of classroom work (Mayer, 1999; Spillane and Zeuli, 1999), unfortunately, few researchers have frequently examined whether teachers truly understand these terms and interpreted them accurately.

In another study, Engel and colleagues (2013) examined the relationship between mathematics instructional content and student knowledge in kindergarten by using the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) data which followed a nationally representative sample of children who were in kindergarten in the 1998–1999 school year through eighth grade (Engel et.al., 2013). The ECLS-K mathematics achievement test measured four mathematics proficiency levels of kindergarten children, including: (a) numbers, shapes, and counting to ten; (b) counting beyond ten, patterns, and relative size; (c) ordinality and sequence; and (d) addition and subtraction. In addition to child data, as part of ECLS-K, teachers reported on their classroom activities and their content by filling out surveys. Particularly, teachers were asked to report how many times they implement specific activities with their students on a daily basis (e.g., Twice a day we counted out loud). Results of the study revealed that children benefited more from being exposed to advanced math content compared to being exposed to more basic math content.

Even though this kind of data, gathered through surveys on mathematics teaching practices, can provide useful descriptive data, such as frequency of instructional grouping choices made by the teacher, most frequently taught mathematical content strands, or materials that teachers are most likely to use to support young children's mathematical learning (Engel et. al., 2013; Hausken & Rathbun, 2004; Porter, 2002), some important aspects of instruction are still missing, such as cognitive demand of the activity, the kind of feedback teachers provide throughout the lesson (Walkowiak, et.al., 2013). Therefore, the relationship between mathematics instructional practice and student achievement may not be fully captured by using these measures.

To sum up, the survey data on instructional practices provide some evidence regarding the frequency of particular approaches and math activities that happen in early childhood classrooms and provide researchers with some insights about what teachers believe they are doing or intend to do while teaching mathematics, but this evidence is still limited for three main reasons. First, what is reported can be subject to misrepresentations of the "assessed" classroom practices because they are only estimates and there is no other source of information that would prove or disprove what is reported by the teacher. Second, surveys are subject to misinterpretation, depending on how questions are formulated and asked. For example, the term "cognitive demand" may represent different things to different subjects, and have its own meaning to each individual respondent. Third, even if surveys measure the instructional practice constructs they are designed to measure, the relationship between them and student achievement may not be fully captured by only using these assessments. Certainly, more research is needed in order to make sure the use of survey methods in measuring teaching practices in mathematics produces more reliable and valid data and to investigate why errors arise when teachers are reporting what they think they teach versus what they actually teach.

Conducting Observations to Measure Mathematics Teaching Quality

Classroom observations are one of the most common forms of teaching evaluation which involves an observer recording what students and teachers do and activities in which they are engaged during a given time interval (e.g., lesson) (Johnson & Christensen, 2012). They are popular among many researchers because they can potentially provide more objective measurement, rich and detailed data, and firsthand accounts of the phenomena being observed.

In recent years, more and more scholars have begun to utilize observation methods to examine the current status of mathematics teaching and its quality (Boston, 2012; Ball & Rowan, 2004; Smith, Lee, & Newmann, 2001; Weiss, Pasley, Smith, Banilower, & Heck, 2003; Walkowiak, et. al., 2013). For example, the Inside the Classroom study by Horizon Research observed and analyzed a representative sample of more than 350 mathematics lessons in order to understand and assess mathematics instructional quality in the U.S. across grades K-12 (Weiss, Pasley, Smith, Banilower, & Heck, 2003). Observers were asked to rate several individual indicators, such as “The design of the lesson reflected careful planning and organization,” and “The instructional strategies and activities used in this lesson reflected attention to students’ experience, preparedness, prior knowledge, and/or learning styles.” Each indicator was rated on a scale 1 to 5, with “1” designating “poor” and “5” designating “excellent.” Based on analysis of these observations, researchers were able to detect varying degrees of mathematics teaching quality among observed teachers. The findings of the study suggested that only 15 percent of the observed lessons were identified as high in quality while 27 percent were medium and 59 percent were low (Weiss, et. al., 2003).

Observations of mathematics lessons can also be used to reveal what makes certain teaching strategies more effective than others and allow researchers to examine the effects of implementation of instruction on students’ achievement (Boston, 2012; Ball & Rowan, 2004). For example, Klibanoff, Levine, Huttenlocher, Vasilyeva & Hedges (2006) observed and audio-taped math speeches of a total of 26 head teachers and 198 children from 13 preschools and day-care centers in Chicago. One of the important results of their study was the finding that mathematizing children's daily experiences and explaining them in explicit math language was significantly associated with their students' math knowledge growth. Doabler and colleague investigated the extent to which explicit instruction in early mathematics instruction is critical for improving kindergarten students' mathematics achievement and found a similar result to those of Klibanoff (Doabler & Fien, 2013). A total of 379 observations were conducted in 129 kindergarten classrooms, involving approximately 2,700 students from 46 schools. Results indicated that providing explicit mathematics instruction was significantly correlated with student students' math achievement.

In another study, the TIMSS 1999 Video Study, the researchers randomly collected nationally representative samples of 8th-grade lessons in mathematics and science. These lessons were videotaped in the U.S. and also in a number of countries in Asia and Europe. Analysis of these videotaped lessons revealed that high achieving countries, such as Hong Kong, Japan and Netherlands, teach 8th grade mathematics very differently compared to low achieving countries such as the U.S. (Stigler & Hiebert, 2004). For example, even though every country showed variations in the kinds of problems that they emphasized during math lessons, there was one important similarity among the high achieving countries. In these countries, in fifty percent of the problems presented to the students, teachers drew students' attention to the connections and relationships between the problems. In comparison to their high achieving international peers, U.S. students were asked or prompted to explore and discuss mathematical relationships between problems less than one percent of the time. By observing and coding varying mathematics instruction across different countries, these researchers were able to identify specific teacher practices that are associated with student achievement and students' engagement with high level cognitive tasks in mathematics. Furthermore, based on these results, authors suggested that U.S. teachers need to make accommodations in their mathematics instruction by noting that

> the results suggest that some time should be devoted to practicing skills and some time devoted to developing understanding. U.S. teachers already provide practice on skills. This now needs to be balanced with solving challenging problems and discussing the relationships that can be constructed among the mathematical facts, procedures, and ideas. When working on these problems, teachers must learn how to avoid stepping in and giving the answers, and instead provide students with opportunities to think more deeply about mathematical concepts and then discuss these concepts or relationships with the students. (Stigler & Hiebert, 2004, p.13)

Observations of mathematics instruction also matter because beyond serving as a monitoring tool, they have the potential to improve the instruction of individual teachers in early mathematics. By looking inside the classroom practices and detecting where each teacher needs support, researchers and practitioners can generate and provide helpful and timely feedback for teachers to further their professional learning and practice in early mathematics (Chen & Cerezci, 2014; Vygotsky, 1978; Walkowiak, et. al., 2013). Feedback from individual teaching profiles derived from systematic observations has been found to help teachers understand their own strengths and weaknesses, and has consequently enabled them to significantly improve their instruction in early mathematics (Chen & Cerezci, 2014).

Despite their benefits, using observation methods can also pose several limitations. For example, when measuring quality of mathematics teaching through classroom observations, using valid and reliable observation instruments are important. Equally crucial are well-trained and calibrated observers who would use these tools in accordance with the protocol of the tool. Rater reliability is also another important concern. Even though the field has made some progress in developing methods to train and calibrate evaluators to ensure more consistent ratings, there is no assurance that a given research project actually employs these methods. When that is the case, the utility and credibility of the protocols themselves can be compromised. Furthermore, without systematic use of standardized, reliable, and validated observational tools, the value of any observations and the feedback they provide to teachers is limited and even questionable. Therefore, it is important to choose which observation tools to utilize thoughtfully and to administer them in ways that minimize any limitations. Thus, when using observations to measure teaching quality in general and mathematics teaching quality specifically, researchers need to use well-validated instruments and train and calibrate raters in order to obtain the most accurate results.

In summation, previous attempts to measure quality of early mathematics instruction have yielded limited results. Research suggests that many of the more commonly used methods of data collection intended to gather data about the nature of instruction (i.e., surveys) produced mixed results in measuring mathematics teaching quality and understanding its effects on students' academic achievement (Ball & Rowan, 2004; Rowan, Correnti, & Miller, 2002; Walkowiak et. al., 2013). Consistent evidence suggests that in order to understand mathematics teaching quality and its effects on students' mathematics achievement, research needs to focus on observing how teachers instruct and provide opportunities for students to learn a significant amount of mathematics. In comparison to survey methods, observations can be considered the most direct way to measure teaching practice because the rater can see the full dynamic of the classroom. Observation results have been modestly to moderately linked to student achievement more consistently than teacher surveys. Furthermore, unlike surveys, observations of early math instruction in early childhood settings can reveal common pitfalls in early mathematics teaching and also highlight specific teaching practices that support and improve student outcomes in mathematics achievement. Thus, developing and using valid and reliable tools that measure mathematics instructional quality in the early years is highly important and even vital. Using such tools would help researchers and educators define what entails quality of early mathematics teaching and create a framework for how to provide high quality mathematics instruction to all children.

An Analysis of Observation Instruments that Measure the Quality of Early Mathematics Teaching

Observation instruments are increasingly being utilized to document the variation in mathematics teaching quality and its multiple aspects. They can allow researchers to document the finer-grained interactions between teachers and students that may have unique and direct effects on how well teachers teach early mathematics lessons and how well children's early mathematics development is supported. Recently, many researchers have started to design and validate a variety of observation instruments which rely on trained observers' interpretations and descriptions of specifics of the mathematics lesson observed (e.g., setting, lesson design, content, and delivery) (Pianta & Hamre, 2009;

Walkowiak, et.al., 2013; Boston & Wolf, 2006; Piburn, Sawada, Falconer, Turley, Benford, & Bloom, 2000).

The option of designing an observation tool can be very appealing, but it can also be a great undertaking for any researcher or research project. It is appealing because designing an observation tool, or any kind of research instrument for that matter, allows researchers to design a measurement that reflects their projects' specific research objectives (e.g., improving mathematics education quality in elementary grades) and conceptual framework (e.g., Vygotsky's theory of development). However, it can also be a very complex task because researchers, who are developing their particular observation measurements, as with any other measurement tool, need to identify accurate, coherent, reliable and valid indicators that operationalize the constructs (e.g., quality of mathematics instruction) that the tool intends to measure. Unfortunately, not every researcher or project has adequate enough funding and expertise to design measures that are theoretically founded and rely on (?) well-defined indicators which are proven to be valid and reliable.

In the process of investigating available observation instruments that are specifically developed to measure mathematics teaching quality in early childhood, four tools emerge as the mostly frequently recommended and used tools: *Reformed Teaching Observation Protocol* (RTOP) (Piburn et. al., 2000), *Mathematical Quality of Instruction* (MQI) (Learning Mathematics for Teaching, 2014), *Classroom Observation Student-Teacher Interactions-Mathematics* (COSTI-M) (Doabler, Baker et al., in press), and *Classroom Observation of Early Mathematics-Environment and Teaching* (COEMET)

(Sarama & Clements, 2007). Below each tool will be briefly introduced. When the information is available, discussion will focus on the conceptual framework, identified mathematics constructs, defined indicators, targeted grade-levels, and psychometric properties. The pros and cons of each instrument, including its applicability to preschool and kindergarten settings will be addressed as well.

Reformed Teaching Observation Protocol (RTOP) is an observation tool that was initially designed for university-level classes to measure the extent to which instruction and interactions between the teacher and students are standards or reform-based in university level classes (Piburn, et. al., 2000). Recently, several researchers have used this tool to rate the quality of mathematics lessons in kindergarten through university (Walkowiak, et. al., 2013).

RTOP consists of 25 items grouped under five subscales: (1) Lesson Design and Implementation, (2) Content- Propositional Pedagogic Knowledge, (3) Content- Procedural Knowledge, (4) Classroom Culture- Communicative Interactions, and (5) Classroom Culture- Student/Teacher Relationships. Contextual background and a brief description of the lesson is also recorded. Each RTOP item listed under each subscale is coded on a Likert scale of 0-4, with zero indicating the item “never occurred” to four indicating the item is “very descriptive” of the instruction (see Table 3).

Table 3. Sample Subscale from RTOP Coding Guide: Lesson Design and Implementation. Adapted with permission from Piburn, et. al., 2000 (redrawn).

Item	Never Occurred				Very Descriptive
1. Instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	0	1	2	3	4
2. The lesson was designed to engage students as members of a learning community.	0	1	2	3	4
3. In this lesson, student exploration preceded formal presentation.	0	1	2	3	4
4. This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	0	1	2	3	4
5. The focus and direction of the lesson was often determined by ideas originating with students.	0	1	2	3	4

The authors of the RTOP examined construct validity and theoretical integrity of the instrument, by performing a correlational analysis on the five RTOP subscales in which each subscale score is used to predict the total RTOP score. The results revealed that all RTOP subscale scores are good predictors of the total score (all R-squared> .75) and offer strong support for the construct validity of the RTOP.

The authors also examined RTOP's predictive validity by observing mathematics lessons of a total 6 university instructors twice during the fall semester in 1999 and administering pre- and post-tests in mathematics to their students. An average RTOP score was created for each instructor and later correlated with normalized gain scores[1] of their students. Final analysis revealed that all correlations between the RTOP and

[1] Normalized Gain= (Post-test score-Pre-test Score)/(Total Score-Pre-test Score) (Piburn et al., pp.13-14, 2000).

normalized gains of students were .88 or higher indicating valid inferences can be drawn from RTOP scores. Also, inter-rater reliability estimates for RTOP coders have been reported as .954 (Piburn et al., 2000).

RTOP exhibits a couple of desirable characteristics for a measure of mathematics classroom instruction as well as limitations. RTOP is aligned with NCTM Standards (2000). As such, the instrument design is grounded in a deep understanding of mathematics teaching. Further, documentation on its validity and reliability is very thorough and present specific measures of its psychometric properties indicate that it is a reliable and valid tool. However, because it is primarily designed and validated to be used in higher education, RTOP exhibits limited applicability to be used in early grades such as kindergarten (Kilday & Kinzie, 2008) and no applicability in preschool settings. For example, one of the items listed under "Lesson Design and Implementation" dimension is "In this lesson, student exploration preceded formal presentation" (See Table 3). This level of mathematics teaching might not be frequently observed in early childhood classrooms, especially in preschool and kindergarten classroom because "formal presentation" is simply not a developmentally appropriate practice for this age group. Observing math lessons at these grade levels by using this tool, in that sense, might not yield accurate conclusions about the quality of the observed lesson.

Mathematical Quality of Instruction (MQI) 4-point version is designed to evaluate the quality of video-taped mathematics instruction and content by rating the teacher-student, teacher-content, and student-content interactions in K through 9th grade classrooms (Learning Mathematics for Teaching, 2014). The conceptual framework for

MQI suggests that there are four dimensions of mathematics teaching: (1) richness of mathematics, (2) working with students and mathematics, (3) errors and imprecision, and (4) common core aligned student practices. These dimensions are grouped under "Segment Codes." Each previously videotaped lesson is divided into multiple segments, and each lasts approximately five to seven-and-a-half-minutes for scoring. Raters assign each segment a score for each of the four MQI elements based on a 4-point scale (1 being not present and 4 being high quality) (see Table 4). Apart from "Segment Codes," MQI 4-point also asks observers to assign "Whole Lesson Codes" based on multiple indicators (i.e., lesson time is used efficiently, lesson is mathematically dense, students are engaged, lesson contains rich mathematics, teacher attends to and remediates student difficulty, teacher uses student ideas, mathematics is clear and not distorted, tasks and activities develop mathematics, lesson contains common core aligned student practices and whole-lesson mathematical quality of instruction).

Table 4. Sample Dimension from MQI Segment Codes: Richness of the Mathematics. Adapted with permission from LMT, 2014 (redrawn).

Indicator	Not Present	Low	Mid	High
Overall Richness of the Mathematics	Elements of richness are present but are all incorrect OR Elements of rich mathematics are not present	Elements of rich mathematics are minimally present. Note that there may be isolated Mid scores in the codes of this dimension	Elements of rich mathematics are more than minimally present but the overall richness of the segment does not rise to the level of a High. For example, a segment may be characterized by some Mid scores in the codes of this dimension or by an isolated High along with substantial procedural focus, etc.	Elements of rich mathematics are present, and either: There is a combination of elements that together saturate the segment with rich mathematics either through meaning or mathematical practices. OR There is truly outstanding performance in one or more of the elements.

Differently from how "Segment Codes" are rated, raters assign a score to these indicators based on a 5-point scale (1 being not at all true of this lesson and 5 being very true of this lesson) instead of 4-point scale (see Table 5).

Table 5. Sample dimension from MQI Whole Lesson Codes: Students are Engaged Adapted with permission from LMT, 2014 (redrawn).

Not at all true of this lesson 1	2	(Default Score) 3	4	Very true of this lesson 5
Students are not engaged with the lesson; many are off task for all or part of the lesson.		Students complete the requests made by the teacher, but do not appear eager to participate.		Students are eager to participate the lesson. They raise their hands or call out answers. Most students are engaged in this fashion.

MQI manual does not provide any explanation for how each segment scores are calculated and analyzed to assign an overall score for particular dimension. In order to get an accurate picture of the quality of mathematics teaching practices, the authors suggest collecting a total of 3 observations per teacher. Two raters working independently to score each lesson and scores are averaged across lessons to assign the teacher a composite score per observation.

Inter-rater reliability was reported as 80% for richness of mathematics dimension, 68% for working with students dimension, 75% for errors dimension, 82% for students participation dimension and 77% for overall MQI score (Learning Mathematics for Teaching, 2014).

As an observation tool to assess mathematics teaching quality, MQI exhibits strengths as well as weaknesses. Firstly, MQI manual provides rich and detailed descriptions of each MQI dimension and its indicators. This level of explanation would allow researchers and assessors to understand the tool very thoroughly. In contrast to the detailed description of the tool, the authors fail to provide the same level of detail in describing the theoretical bases of MQI, which leaves the reader and user wondering how MQI dimensions and indicators hold together to reflect the quality of mathematics teaching and learning. Also, even though the authors reported high level of inter-rater reliability estimates, they failed to provide any information on how assigned scores are calculated to determine the overall MQI scores and quality of mathematics instruction. Further, even though the authors of the MQI stated that this tool is developed to be used in K through 9th grade, it is actually used in higher grades (Hill, et. al., 2008) similar to RTOP, and the dimensions and indicators listed are more in line with elementary mathematics content rather than kindergarten.

Classroom Observation Student-Teacher Interactions—Mathematics (COSTI-M) is an observational tool designed to document the frequency of explicit instructional interactions that occur between teachers and their students during kindergarten mathematics instruction (Doabler, Baker et al., in press). The COSTI-M includes two sections, the Context Codes and the Instructional Interaction Codes. While the Context Codes section documents (a) the duration of the instructional time, (b) content of the mathematics activity, and (c) the instructional format. The Instructional Interaction Codes collect data on (a) teacher demonstrations, (b) teacher-provided academic feedback, (c) group responses, (d) individual responses, (e) student errors, and (f) other forms of student responses. Observers record these behaviors as they occur. At the end of the observation total number of observed behaviors is calculated (see Table 6).

Table 6. Sample dimension from COSTI-M Coding Guide: Individual Response. Adapted with permission from Doabler, Baker et al., in press (redrawn).

Indicators
A single student produces an answer
The answer is preceded by a teacher-posed question or request
A student's answer immediately follows a teacher-posed question or request
An individual student other than the one identified by the teacher responds ("callout")

Reliability analysis of the COSTI-M items suggests that individual response opportunities and academic feedback dimensions were modestly stable over time (intra-class correlation coefficients [ICC]s = .34 and .35, respectively) Stability ICCs for other COSTI-M behaviors range from .13 to .19. Reported inter-observer reliability ICCs for the COSTI-M were .67 for teacher models, .92 for group responses, .95 for individual responses, .91 for other forms of responses, .84 for errors, and .90 for feedback (Doabler, Baker and et al., in press).

In order to document the predictive validity of the COSTI-M, the authors used a dataset obtained from *Early Learning in Mathematics* (ELM; Clarke et al., 2011). This efficacy trial included 129 kindergarten classrooms from 7 school districts and 46 schools in Oregon and Texas. The sample included 129 teachers and 2,103 students at pretest and 2,270 students at posttest. Results provided preliminary evidence for the COSTI-M's predictive validity with the Test of Early Mathematics Ability-3rd Edition (TEMA-3), a broad, standardized measure of mathematics achievement (p=.004, pseudo-= .08), and a battery of early mathematics curriculum based measures (p=.017, pseudo-= .05; see Doabler, Baker, et al., in press).

COSTI-M is specifically designed to be used in kindergarten classrooms and reported to be valid and reliable research tool. Further, authors of COSTI-M also documented significant correlations with direct assessments' of math outcomes, providing preliminary evidence for the validity of the COSTI-M as an observation measure. Despite these desirable characteristics, COSTI-M also exhibits important limitations. First, the COSTI-M manual does not provide any detailed descriptions of COSTI-M items and how they are connected conceptually and theoretically in terms of measuring mathematics teaching quality and learning in kindergarten classrooms. This kind of lack of clarity in item description makes it harder for assessors and other researchers to understand the tool thoroughly. Second, even though the authors present this tool as an observation instrument, indicators listed under COSTI-M dimensions asked the raters to document the frequency of certain behaviors, not necessarily their quality (see Figure 3).

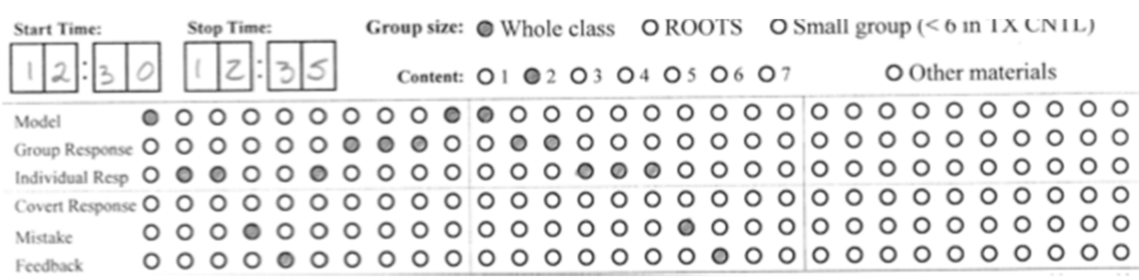

Figure 3. COSTI-M Coding sequence example.

Classroom Observation of Early Mathematics-Environment and Teaching (COEMET) is an observation tool specifically designed to measure the preschool instructional environment in mathematics (Sarama & Clements, 2007). COEMET consists of a total of 28 items that are grouped under two main sections: Classroom

Culture (CC) and Specific Math Activity (SMA). The Classroom Culture portion of COEMET is intended to measure the general classroom environment throughout the observation and only completed once by reflecting on overall evidence gathered from the entire observation. This section consists of a total of 9 items. For each item, observers are asked to rate whether or not they agree with the statement listed on that particular item (e.g., “The teacher showed curiosity about and/or enthusiasm for math ideas and/or connection to other ideas or real world situations”). The authors stated that this section is developed to yield information on how the teacher: interacts with students; utilizes teachable math moments; and displays math in the classroom.

The Specific Math Activity (SMA) portion is intended to measure the quality of intentional mathematics activities (from 0 to 12) and interactions involving the teacher and one or more children. Each SMA is rated based on a total of 19 items that are grouped under seven dimensions: (1) mathematics focus, (2) organization, teaching approaches and interactions, (3) expectations, (4) eliciting children’s solution methods, (5) supporting children’s conceptual understanding, (6) extending children’s mathematical thinking, and (7) assessment and instructional adjustment (see Table 7).

Observers complete a separate SMA each time a math activity occurs that lasts more than 30 seconds. Apart from rating COEMET items for each activity, observers also take note of the duration of the activity and instructional grouping choices that are made throughout the activity. Similar to CC sections, most items in SMA are coded on a Likert scale from Strongly Disagree to Strongly Agree (e.g., “The teacher displayed an understanding of mathematics concepts”).

Table 7. Sample dimension from COEMET—SMA Coding Guide: Mathematical Focus. Adapted with permission from Sarama & Clements, 2007 (redrawn).

Statements	Indicators	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
10. The teacher displayed an understanding of mathematics concepts.						
11. The mathematical content was appropriate for the developmental levels of the children in this class.	- Used task at the level of difficulty consistent with children's level of thinking and learning. - Used tasks in sequence corresponding to children's growing level of thinking.					

In terms of the frequency, assessors record the approximate percentage of occurrences for which the statement is true (e.g., 0%, 1-25%, 26-50%, 51-75%, and 76-100%). In order to complete all parts of the COEMET, assessors spend no less than a half-day in the classroom. High levels of inter-rater reliability (.88), internal consistency (Cronbach's alpha .94) and Rasch model reliability (.96) have been reported (Clements & Sarama, 2008).

COEMET is primarily developed to be used in preschool settings and exhibits several desirable characteristics in terms of its theoretical base, reported psychometric estimates, and targeted population. Firstly, COEMET's framework and goals are specifically aligned with NCTM Standards (2000). Secondly, documentation on COEMET's reliability demonstrated acceptable internal consistency. Reliability and inter-rater reliability levels meet the standards used by the experts in the field (values greater than .70 and .85 respectively). Despite its applicability in preschool settings and other desirable characteristics, COEMET also exhibits several weaknesses. Firstly, COEMET's dimensions and indicators under each dimension are not thoroughly explained by the authors. This kind of lack of clarity in item description might increase subjective interpretation of what each item means and how it needs to be measured by the assessors. Secondly, there is no report on the tool's validity that indicates what COEMET is actually measuring what it intends to measure. Further, tools such as COEMET are difficult to be used in large-scale studies due to the nature of the tool (observers are asked to spend approximately 4.5 hours to observe).

The review of currently available observation tools developed to measure instructional quality in early mathematics revealed that while there are a variety of tools developed to measure mathematics teaching quality in kindergarten through higher education (e.g., M-Scan, RTOP, MQI, IQA), there are only few measures available and specifically developed to be used in kindergarten (e.g., COSTI-M) and preschool (e.g., COEMET) classrooms, a noteworthy gap. Even though most of available observation tools could potentially provide useful information on many facets of mathematics instruction, each exhibited varying degrees of strengths and weaknesses for a measure of quality of mathematics' instruction and limited applicability in kindergarten and preschool settings (see Table 8).

Even though COEMET and COSTI-M were primarily designed to be used pre-kindergarten and kindergarten settings, lack of conceptual framework, limited reporting their inter-rater reliability and validity estimates, and how the indicators of quality of mathematics teaching explained and measured make them less desirable for research purposes.

Table 8. Observation Instruments Designed to Measure Mathematics Teaching Quality in Early Childhood Settings

Measure	Grade Level	Framework	Constructs Measured	Reliability	Validity
RTOP	K-University	NCTM Standards (2000) National Science Education Standards (1996)	- Lesson design and implementation - Content - Classroom culture	Inter-rater: = .954	Construct: all R-squared's > .75 Predictive: all correlations between RTOP and normalized gains > .87
MQI	K-9th	NCTM Standards (2000)	- Richness of mathematics - Working with students and mathematics - Errors and imprecision - Common core aligned student practices	Inter-rater: = .77	No Report
COSTI-M	K	No Report	- Teacher demonstrations - Teacher-provided academic feedback - Group responses - Individual responses - Student errors - Other forms of student responses	Inter-rater: =.67	Preliminary evidence
COEMET	Pre-K	NCTM Standards (2000)	- Mathematics focus - Organization, teaching approaches and interactions - Expectations - Eliciting children's solution methods - Supporting children's conceptual understanding - Extending children's mathematical thinking - Assessment and instructional adjustment	Inter-rater: = .88 Internal consistency: α>.94 Rasch model reliability is .96	No Report

As policy makers and researchers focus more on the importance quality of early mathematics teaching and its implications on students' mathematics learning, new measures will be developed to measure the quality of mathematics instruction in preschool and kindergarten settings. In order to maximize the usefulness of these tools, the researchers not only need to conceptualize each tool's framework based on the latest recommendations and standards in the field, but also need to define the constructs that the tool intends to measure and describe its indicators clearly and conceptually. Improving the student outcomes in mathematics is one of the main reasons researchers have developed various tools to quantify the quality of instructional quality in early mathematics. In order to maximize children's learning in mathematics, we need to be able to develop tools that will help us understand the quality of instruction and identify those teaching characteristics that promotes or hinders students' learning in mathematics.

For the field to progress, it is important that measurements of quality of instruction in mathematics are also methodologically well-designed. Any instrument that is designed to be used widely across different settings should be able to produce the same results over time and/or across raters (i.e., reliability) and measure what it intends to measure (i.e., validity). Especially, the predictive validity heavily depends on how well the tools are developed and validated. Accomplishing this is not an easy task. It is challenging because developing and validating measurement tools can be very costly and might require extensive support financially and scholarly. Reliable and valid early mathematics observation tools would allow researchers not only examine the relationship among many aspects of quality early mathematics teaching and student outcomes but also provide teachers with targeted feedback on ways to improve their delivery of early mathematics instruction. To that end, there is a need for a reliable observational measure with a strong theoretical framework which would focus on studying the mathematics instructional interactions that may develop early mathematics skills in preschool and kindergarten settings and further can be used to guide early math instruction.

Observing Early Mathematics Teaching with High Impact Strategies in Early Mathematics (HIS-EM) Measure

Quality of instruction and students' instructional experiences in early mathematics—that is, their experiences learning early math concepts through instruction— lay the foundation for the formal systems of math that will be taught later in school. Despite its importance, our knowledge about what constitutes effective instruction, how it looks in practice, and how to quantify it is quite limited (Ball & Rowan, 2004; Brenneman, et. al., 2011). A review of currently available observation tools intended to measure quality of early mathematics teaching revealed a pressing need for conceptually-founded, reliable and valid observation tools. In response to the field's need, Early Math Collaborative at Erikson Institute developed the High Impact Strategies in Early Mathematics (HIS-EM) observation instrument to shed some light on how math-related instructional interactions that occur during early mathematics lessons look like across settings and how different instructional practices affect achievement in mathematics. This section describes HIS-EM's conceptual model and how the HIS-EM tool is positioned in in the context of current literature.

The Conceptual Model for HIS-EM

The HIS-EM conceptual model presented in this section describes and outlines the observable components of teaching that can deepen young children's understanding of fundamental math concepts and ideas. It assumes that the best early math teaching will be based on, and therefore reflect teachers' pedagogical content knowledge for early mathematics. Specifically, the conceptual model of HIS-EM uses the pedagogical content knowledge framework (PCK) developed by Shulman (1986) as a model to determine the quality of teaching practices in early mathematics. Because the HIS-EM is centrally influenced by a PCK framework, a closer look at it and its components follows.

Pedagogical Content Knowledge (PCK)

Lee Shulman (1986) advanced thinking about teaching by introducing the idea of pedagogical content knowledge (PCK). It is a notion which represents the blending of content and pedagogy into an understanding of how to organize and deliver the subject-matter in order to make it comprehensible to others. More specifically, he defined PCK as including

> the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations [...] [it] also includes an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of the most frequently taught topics and lessons (Shulman, 1986, p.9).

According to Shulman, PCK is different from the knowledge of the disciplinary expert (e.g., mathematician) and from the general pedagogical knowledge shared by professionals and teachers across various disciplines (e.g., general child development knowledge). Rather, it is a distinct body of knowledge specific to teaching and "how to represent specific subject matter topics and issues appropriate to the diverse abilities and interest of learner" (Shulman & Grosman, 1988, p.9). Further, he states that in order to successfully blend content and pedagogy, teachers need to embody “the aspects of content most germane to its teachability” (Shulman, 1986, p.9). In this sense, what lies at the heart of PCK is how the subject-matter is transformed for teaching; how this transformation occurs can influence the quality of teaching practice because it is closely related to “the ways of representing and formulating the subject that make it comprehensible to others” (Shulman, 1987, p. 9).

The introduction of PCK has driven more research into teacher knowledge because it has been perceived as a useful notion which blends the traditionally separated knowledge bases of content and pedagogy. Since its introduction, many empirical studies have been conducted on the essential components of PCK and the role of PCK in teaching different subjects like mathematics (Lee, Brown, Luft, & Roehrig, 2007). As a result, regardless of its pre-eminence, Shulman’s notion of PCK has been refined (Graeber & Tirosh, 2008) and expanded by a number of researchers to better understand its components and their effects on quality teaching and student learning in mathematics (Marks; 1990; Krauss, Baumert & Bloom, 2008; Ball and Bass 2000; Hill, Ball, & Schilling, 2008; Hill, Schilling, & Ball, 2004).

Knowledge and Practice in Early Mathematics

One important goal of research in mathematics education is to identify the teaching practices that predict students’ achievement. In the literature, pedagogical content knowledge is one of the characteristics that keep emerging as important determinant of instructional quality that affect students’ learning outcomes (Bransford, Darling-Hammond, & LePage, 2005; Grossman & McDonald, 2008; Hiebert, Morris, Berk, & Jansen, 2007). For instance, in their study Baumert and colleagues (2010) defined PCK as *knowledge of mathematics tasks*—teachers' ability to identify multiple solution paths; *students' thinking*—ability to recognize students' misconceptions, difficulties, and solution strategies; and *multiple representations*— teachers' knowledge of different representations and explanations of standard mathematics problems (p. 149). Based on this definition, they analyzed Grade 10 mathematics teachers' PCK through open-ended questionnaires. Results revealed a significant positive effect of teachers' PCK on instructional quality (assessed by means of students' ratings on teachers' quality of adaptive explanations, responses to questions, pacing and teacher-student interaction) and on student outcomes in mathematics (explaining 39% of variance in students' learning gains over a year) (Baumert, Kunter, Blum, Brunner, Voss, Jordan, Kusmann, Krass, Neubrand, & Tsai, 2010).

Ball and colleagues (Ball and Bass 2003; Hill, Ball, & Schilling, 2004; Hill, Schilling, & Ball, 2008) developed a construct called 'mathematical knowledge for teaching' (MKT) and defined it as being composed of two major categories: subject matter knowledge and PCK. Subject matter knowledge contains: 'common content knowledge' (CKT) referring to mathematical knowledge that a literate general population might have; 'specialized content knowledge' (SCK) referring to specific knowledge for teaching math to specific groups of students; and knowledge at the mathematical horizon which concerns how concepts are introduced over grades.

In this model, the authors characterized PCK as having three components: knowledge of content and students (KCS), knowledge of content and teaching (KCT), and knowledge of curriculum (Ball and Bass 2000; Hill, Ball, & Schilling, 2004; Hill, Schilling, & Ball, 2008) (see Figure 4).

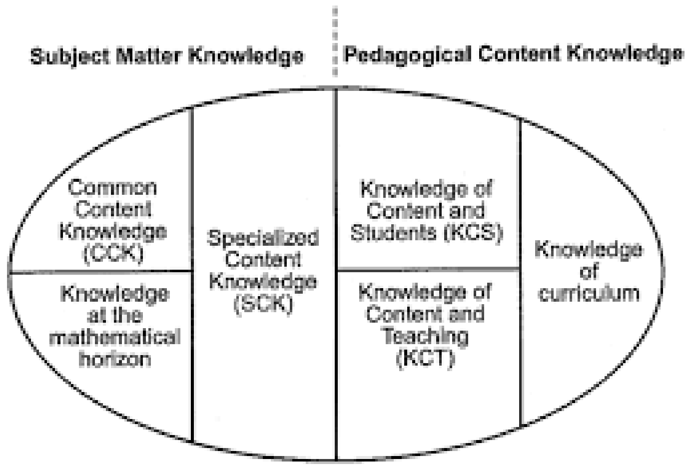

Figure 4. Domains of Mathematical Knowledge for Teaching (MKT)

In one of their studies, Hill and colleagues investigated the relationship between teachers' mathematical knowledge for teaching and quality of instruction in grades K through 8th (Hill, Blunk, Charalambous, Lewia, Phelps, Sleep, & Ball, 2008). Results suggested that there is a strong relationship between what teachers know about mathematics, how they know it quality of mathematics instruction. They also found that there are a number of important factors that mediate this relationship (e.g., beliefs about mathematics).

Kersting and colleagues (2010) defined PCK as including understanding of the content, understanding of students, and understanding of pedagogy (Kersting, Givvin, Sotelo, & Stigler, 2010). They used this definition to investigate elementary teachers' PCK in mathematics in relation to student outcomes. The researchers asked teachers to watch several classroom teaching video clips and to describe "how the teacher and students interacted around the mathematical content” (p.174). Results suggested that there is a positive association between teachers’ PCK scores and students outcomes.

Even though there are some empirical studies investigating the relationship between PCK and instruction in primary mathematics education, there is a dearth of large-scale studies in early mathematics in preschool and kindergarten. Among these few investigations, McCray and Chen (2012) studied preschool teachers’ PCK in relation to instruction and student outcomes. According to these researchers PCK for preschool teachers includes “an understanding of the foundational concepts of mathematical content, combined with the skill to closely observe children’s play, discern their likely thinking, and provide language that points out embedded mathematics” (McCray & Chen, 2012, p. 304). In their study with 22 preschool teachers with 113 students in Head Start preschools, the researchers used an applied scenario-based interview to explore preschool teachers’ PCK in mathematics. They also gathered data on mathematical language used during the instruction and students’ performance on standard tests. Despite the statistical limitations of the study (i.e., small sample size), the researchers were able to find a significant association between preschool teachers’ PCK and quality teaching.

In the research literature on mathematics teaching and learning, there is a shared understanding that content-specific knowledge, general pedagogical knowledge and skills, and knowledge of learners are important determinants of instructional quality that affect students’ learning gains. Effective teaching entails an integration of these different knowledge domains (Park & Oliver, 2008). Byrne (1983) explains this point well by stating

> It is surely plausible to suggest that insofar as a teacher's knowledge provides the basis for his or her effectiveness, the most relevant knowledge will be that which concerns the particular topic being taught and the relevant pedagogical strategies for teaching it to the particular types of pupils to whom it will be taught. If the teacher is to teach fractions, then it is knowledge of fractions and perhaps of closely associated topics which is of major importance....Similarly, knowledge of teaching strategies relevant to teaching fractions will be important (p. 14).

What these domains of knowledge look and sound like in practice has not been well-translated into models for understanding the quality of instruction in early mathematics, nor have the underlying skills and understandings required for these domains of knowledge been well articulated. Investigating the quality of instruction in mathematics in early childhood classrooms— that is, searching for a particular pattern of instructional variables and conditions in mathematics teaching that influence student achievement in mathematics and meets the demands of the discipline, goals of instruction, and range of students' learning needs in mathematics— requires a model. Thus, HIS-EM, as a measure of practice, is grounded in the belief that the evidence of PCK in the actions of teachers during math lessons is likely to be the strong indicator of children's learning.

Developing a Conceptual Model for HIS-EM based on PCK

HIS-EM proposes that the interplay between teachers' content knowledge in mathematics, knowledge of students and learning, and knowledge of how to teach mathematics effectively during the course of early mathematics lessons can be observed and will reflect the quality of mathematics instruction provided. Therefore, HIS-EM endeavors to outline the observable characteristics of quality early mathematics teaching by gathering evidence in three areas.

First, it identifies teacher actions indicative of knowledge of foundational mathematics concepts. Second, it requires teachers to be familiar with young children's learning in mathematics and it is built on the notion of learning as developmental progression and instruction should be developmentally appropriate. Third, it focuses on the kinds of instructional practices needed to be used in order to engage young children in developmentally appropriate and meaningful learning experiences during mathematics instruction.

Thus, HIS-EM seeks to illustrate how content knowledge, knowledge of development of young children and their math understanding and use of appropriate and effective instructional strategies in teaching mathematics are interwoven in practice in order to provide quality mathematics instruction and learning experiences in early years (see Figure 5).

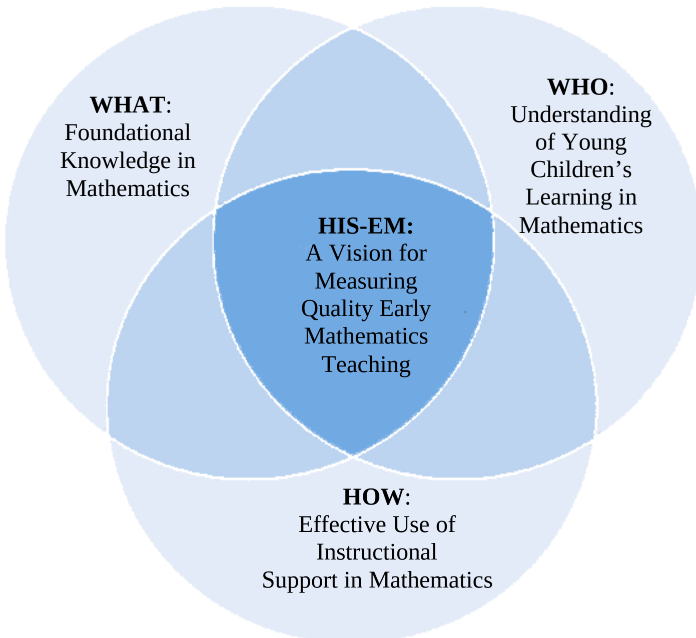

Figure 5. Conceptual Model of HIS-EM for Observing Early Mathematics' Instruction.

What: Foundational Knowledge in Mathematics. *What* refers to the degree to which observed practice incorporates deep knowledge of foundational mathematics concepts in teaching early mathematics to young children. It requires teachers to have comprehensive and in-depth knowledge of early mathematics; and competence in representation and manipulation of this knowledge during instruction in a way that allows all students to engage in conceptual math learning (Battista, 1999; Cohen, Raudenbush & Ball, 2003; Shouse, 2001).

HIS-EM claims quality math instruction that focuses on fostering deep knowledge of foundational math concepts can be observed by focusing on: (1) how well and clearly the teacher emphasizes the learning objectives, which reflect important learning and conceptual understanding in mathematics, (2) in what ways teacher promotes the use of multiple mathematical representations (e.g., mathematical language, tools and models) to illustrate and connect math ideas and concepts accurately and coherently, and (3) the extent to which teacher's mathematical content knowledge is accurate and coherent in a way that allows her to help students generalize their understanding of the key math concepts.

Formulating a clear and conceptual learning objective and making the children aware of it at the beginning of the lesson sets the stage for encouraging active student learning in mathematics. Further, without maintaining the focus on learning objectives throughout the lesson, it is almost impossible to understand what the teacher wants children to focus on and what counts as evidence of students' learning in relation to instructional activities. However, having explicit learning objectives by itself is not enough for quality teaching and student learning.

Quality learning objectives in early mathematics lessons need to emphasize mathematically important and developmentally appropriate learning goals that are linked to *big ideas of early mathematics education*— ideas that connect key mathematical concepts to promote coherent and meaningful mathematics learning (Clements & Sarama, 2009; Erikson Institute's Early Math Collaborative, 2013). National Council of Teachers of Mathematics recommends that early mathematics instruction cover the "big ideas" of mathematics in such areas as number and operations, geometry (shape and space), measurement, and algebra (particularly pattern); within learning contexts that promote problem solving, analysis, and communication (NCTM, 2006). For example, during a lesson with the goal of helping students use their knowledge of 2-dimensional shapes to categorize 3-dimensional shapes by their flat faces, the teacher connects students' learning to one of the big ideas in geometry (i.e., analyzing and comparing attributes of shapes helps one define and classify shapes) and students' prior knowledge in an intentional and meaningful way. Formulating this kind of conceptual learning objective for early mathematics lessons largely depends on teachers' awareness of the diverse, yet connected foundational math concepts.

Building a deeper foundational knowledge in mathematics, however, is more than identifying clear and conceptual learning goals. Utilizing accurate and appropriate mathematical representations is equally essential (Clements & Sarama, 2009; English & Halford, 1995; Ball, 1992). Mathematical representations (e.g., mathematical language, tools and models), when used purposefully, can help students to investigate mathematical concepts and processes and increase students' flexibility of thinking (English & Halford, 1995; Varol & Farron, 2006). However, providing mathematics tools and manipulatives by themselves is not enough, because they do not magically create coherent and conceptual mathematical understanding (Ball, 1992). Rather, they provide concrete ways for students to understand the math topics they have been introduced to. Therefore, it is crucial for teachers to connect mathematical tools and models with mathematical concepts in which students can engage in through instruction (Dufour-Janvier, Bednarz & Belanger, 1987; Carpenter & Lehrer, 1999).

For example, during a lesson about a standard measurement with a foot-long ruler, the teacher may focus not only on correct procedures to use a ruler, but help students make connections between using a ruler and the non-standard forms of measurement they have previously used. Teachers cannot assume that their students will make the desired interpretations from the concrete representation of the ruler to the abstract idea which in this case is *equal partitioning*—dividing a length into equal size units that can be counted.

Further, for early mathematics teaching to build foundational math knowledge, instruction should also focus on fostering conceptual understanding in mathematics. When teachers provide opportunities for students to develop conceptual math knowledge, they can enable students to apply their math knowledge to learn new topics and solve new and unfamiliar problems (e.g., How is dividing a circle into 4 quarters related to telling time on a clock? How about coins we call "quarters"? Why are "quarters" also called "fourths?"). However, instruction cannot lay the foundation for conceptual mathematical learning if the teacher herself does not have deep subject-matter knowledge in early mathematics. Therefore, while observing math instruction, it is crucial to look for instructional interactions that can indicate the degree of the content knowledge the teacher has in early mathematics.

For example, many students have a limited understanding of what defines triangle because the most common example of a triangle is an equilateral triangle with a horizontal base. When the teacher is unaware of this, and does not offer other examples of triangles, she inadvertently reinforces student misconceptions. When students tell her that a given triangle is "upside down," if she does not correct them, the lesson does not lead students to a deeper understanding of what are (and are not) the defining traits of a triangle. When the instruction does not yield a conceptual understanding of the mathematics, the students tend to perceive each math topic in isolation and fail to apply the newly acquired knowledge in different settings or contexts (Sutton & Krueger, 2002; Fennema & Romberg 1999).

Who: Knowledge of Young Children. *Who* represents the degree to which observed practice reflects an understanding of young children's typical developmental growth in mathematics and understanding of individual students' learning needs. It requires teachers to design math learning experiences that are developmentally appropriate both in content and format (Darling-Hammond, 1999; Sarama & Clements, 2004; 2007).

HIS-EM claims quality math instruction that incorporates knowledge of young children can be observed by focusing on: (1) the teacher's awareness and knowledge of the developmental trajectory for different mathematical ideas, (2) the teacher's awareness of and response to students' different needs in mathematics and the degree to which she facilitates students' ability to actively explore and learn at their own pace by monitoring their work and adjusting the lesson, (3) the degree to which the instructional grouping, format and the pace of the lesson are appropriate and productive for the age of the students.

Children follow natural developmental progressions in learning and development. For example, it would be highly unlikely to expect a baby to run before she can even crawl. Similarly, when it comes to teaching mathematics, it is important that early childhood teachers know the order in which math skills and concepts build on one another and how young children typically learn these concepts and skills as they develop.

This kind of knowledge, when used to tailor instruction, can provide teachers with a road map for developmentally appropriate instruction that can optimize teaching and therefore students' learning. For example, a preschooler with some exposure to numbers and counting may correctly determine the number of objects in a small collection. This skill is critical for her to learn because it lays the foundation for more complex understandings such as addition and subtraction. If a preschooler with a limited math understanding is asked the same question while she is struggling to recognize numbers, she would be less likely to make sense of what is being asked of her because she is not developmentally ready to grasp this more advanced math idea.

Ability to identify the knowledge children already possess and expose them to the content beyond their current skills, but still within their range of abilities, is key in helping children acquire new knowledge and move to the next level in their developmental progression in learning mathematics. Conceptual math learning occurs within *the zone of proximal development (ZPD)* of the child—a distance between a child's ability to solve a problem independently and her ability to solve it with just enough support from a more skilled person in the environment (Vygotsky, 1978). The teacher needs to provide scaffolding that entails adjustments of the tasks to fit the child's level of performance and enables the teacher to lead the child in her ZPD to construct higher level of math understanding (Berk & Winsler, 1995; Radziszewska & Rogoff, 1991; Vygotsky, 1978).

For example, a student is trying to figure out how many playing cards she has but keeps losing count. When the teacher asks the student "*Is there a way you could keep* *track of the cards you've already counted*?" the teacher provides assistance that is at the right level to promote student's development without offering too much help. By working with the student and providing appropriate feedback, the teacher can help the student devise a strategy of making piles of five cards and then counting by 5s to get the total.

If the teacher displays no knowledge of individual students' skills and conceptual understanding and the lesson is presented as "one-size-fits all," the lesson would fail to reach most students may learn how to perform math skills incorrectly or they might not be learning anything at all. Therefore, understanding developmental progression in mathematics is also necessary to understand and respond to students' varying learning needs accordingly. For example, one teacher divides her kindergarten class into small groups for Math Centers. During this time, she is able to work with students in small, ability-leveled groups and tailor the lesson to meet their needs. This kind of approach during math instruction will help teachers to generate the information about what students are thinking, how they are reasoning, and how to adjust the lesson, so that she can provide the right level of support and/or challenge (Sutton & Krueger, 2002).

A teacher's knowledge of young children can also be evidenced in what kind of learning formats she utilizes during math instruction. If they are designed intentionally, both small and large group experiences can be used to tailor instruction for children at different developmental levels (Griffin, 2004). Small group activities, for example, could allow children to share their ideas with their peers and model for one another and allow teachers to better understand and support each child's current level of understanding in mathematics. Large group experiences, on the other hand, allow teachers to introduce math concepts and connect them to other areas of the curriculum through different activities which may require the participation of all students.

Furthermore, along with utilizing varied instructional groupings, using multiple modalities (e.g., auditory, visual, kinesthetic) to teach mathematics to young children will help teachers to illustrate different aspects of a math concepts and gain students' interest and their active hands-on participation (Ornstein & Lasley, 2000). For example, during a lesson on position words (e.g., *next to*, *between*, *beside*), the teacher asks students get up and move in the classroom as she uses one of those position words (e.g., stand *between* the tables, go *under* the table and etc.). This kind of approach also allows teachers to help students connect mathematics to their own lives and their surroundings. In particular, in the above example, the teacher helps students realize that math is not an isolated topic but actually it is part of our daily life and it is relevant to our experiences.

How: Effective Use of Instructional Support. *How* represents the degree to which observed practice includes the effective use of mathematics teaching strategies that facilitates young children's mathematical understanding. It requires teachers to interweave the math content and its accompanying pedagogy by planning coherent and conceptual math lessons, engaging children in purposeful mathematical reasoning and inquiry, and fostering a positive disposition towards mathematics (Clements & Sarama, 2008; Larson & Whitin 2010; NAEYC & NCTM, 2010; NRC, 2009).

HIS-EM claims quality math instruction that incorporates effective use of instructional support can be observed by focusing on: (1) how well the teacher selects and prepares a coherent and well-organized math lesson that helps students focus on math concepts, (2) the degree to which the teacher facilitates opportunities for students to construct and make meaning of mathematical ideas and make use of variety of strategies to solve problems and justify thinking, and (3) the degree to which the teacher's attitudes towards math and the teacher's interactions with students foster a sense community in which all student feel welcomed to share their mathematical ideas and contribute to the lesson and classroom discourse.

Deciding on what to teach and in what order are basic components of lesson planning. However, knowing what to teach by itself might not necessarily lead students to a deeper understanding of the mathematical concepts by the end of the lesson. When teachers know: (a) what mathematical concepts they wish children to understand and design the lesson activities based on the stages through which they develop; (b) what materials they are going to use throughout the lesson and prepare them in advance, they have planned coherent and connected math lessons (Bain & Jacobs, 1990; Wall, Nardi, von Minden, & Hoffman, 2002). When instruction is planned to be developmentally appropriate and conceptually coherent, two things can happen: (1) children can be engaged in effective problem solving and thinking, and (2) an environment that can be created that is mathematically empowering.

In order to provide effective instructional support in mathematics, instruction should also be engaging and purposeful. Meaningful problem-solving require mathematical problems in which students realize that there is more than one possible strategy that can be employed before they reach a solution (Geist, 2000). Asking combination of "*what*," "*how*," and "*why*" questions during mathematics lesson will help students make connections between math concepts at a deeper level and prompt students to think about how to describe their ideas mathematically (Sutton & Krueger, 2002).

For example, while teaching a lesson about estimating distances, students are given the opportunity to share their ideas about how to define a reasonable range and then put them to the test. When a child is given a chance to describe her method for solving a problem to her peers and also hear other children's strategies to solve the same problem, the teacher provides opportunities for the children to learn new ways to apply their math knowledge (Siegler, 1995). Unfortunately, teachers more often use "fill in the blank" questions with an emphasis on getting the right answer. "Consistently asking questions to which there is only one right answer fosters a view of learning that is self-limiting--one that looks for simple "right" answers and simple solutions to complex problems, one that relies on authority rather than on rational judgment to find the "right" answer" (Ornstein & Lasley, 2000, p.184).

By creating an environment where students feel comfortable enough to share their beliefs, ask questions, hypothesize, and make mistakes, teachers can empower children in their learning and promote a sense of mathematical learning communities (Ball, 1991). For example, when the teacher says, "*Is there another way you could think about this problem? Remember what you figured yesterday? Do you think the same strategy could work here?*" she interacts with students, and facilitates interactions among students, in such a way that students feel safe to share their ideas and take risks. For example, before commenting on the answer to a problem, the teacher records all solution strategies that students used whether they are "right" or not. Then the students have a chance to discuss and come to agreement about the most reasonable answer. During these discussions, the teacher not only follows her own teaching agenda but also follows student interests, pace and signals. In this kind of environment, the teacher communicates high expectations for all students and encourages them to share their ideas and solutions about given problems as well as to respond to their classmates' solutions (Sherin, 2002). Further, the teacher must encourage participation of the all students in each lesson in order to foster a sense of community in which all students' mathematical ideas are appreciated and mathematical discussion is not dominated by the teacher or a few students (Ornstein & Lasley, 2000).

Overall, The HIS-EM conceptual model intentionally focuses on how the PCK framework proposed by Shulman (1986) might look and sound like in practice and uses the PCK lens to better understand the quality of teaching in early childhood classrooms through observation. The conceptual model for HIS-EM agrees with Shulman (1986) that components of PCK are inevitable parts of effective teaching and for quality mathematics instruction to occur, early childhood teachers need to familiarize themselves with foundational mathematics content, how young children learn in general and specifically in mathematics, and developmentally appropriate teaching strategies to maximize children's mathematics learning and growth. More specifically, based on the key aspects of the PCK framework, HIS-EM model (i.e., What, Who and How) values the purposefully designed math learning opportunities that encourage children explicitly think, talk, and act on real-life experiences and problem in mathematical ways. Ultimately, the quality of mathematics instruction is determined by the degree to which the teacher helps children to interpret foundational mathematical principles conceptually and supports the development of their intuitive knowledge into robust and transferable knowledge in mathematical thinking with developmentally appropriate ways.

HIS-EM and NCTM's Standards and Principles

The *Principles and Standards for School Mathematics*, published by NCTM in 2000, outlines the principles and standards to promote systemic improvement in mathematics education and is the primary model for standards-based mathematics teaching for Pre-K to 12th grade. Based on the PCK framework, HIS-EM's domains of quality mathematics instruction (i.e., "what," "who," and "how"), as operationalized in this research, are also aligned with the definitions of the standards set forth by the NCTM (2000, 2007) in its six principles (i.e., equity, curriculum, teaching, learning, assessment and technology) and five process standards (i.e., problem-solving, reasoning and proof, communication, connections, and representations (see Table 9).

Table 9. Alignment between HIS-EM Domains and NCTM's Standards and Principles

	HIS-EM Domains			
	What	Who	How	Connection
NCTM Principles				
Equity			X	High expectations and strong support for all students
Curriculum	X	X		A coherent curriculum of important mathematics, articulated across grade levels
Teaching	X	X	X	Teachers who understand what students need to learn and challenge and support them
Learning	X			Instruction that connects prior knowledge with new knowledge
Assessment		X	X	Meaningful and intentional assessment that is useful for both teachers and students
Technology	X	X	X	Use of technology that enhances students' mathematics learning
NCTM Standards				
Problem-solving	X	X		Building mathematical knowledge through problem solving
Reasoning and Proof			X	Developing and evaluating mathematical arguments and proofs.
Communication	X		X	Communicating mathematical thinking coherently and clearly to peers, teachers, and others.
Connections	X	X		Recognizing the connections among mathematical ideas and applying mathematics in contexts outside of mathematics.
Representations	X	X	X	Using mathematical representations to organize, record, and communicate mathematical ideas.

Note:
What: Foundational Knowledge in Mathematics
Who: Knowledge of Young Children
How: Effective Use of Instructional Support

The High Impact Strategies in Early Mathematics (HIS-EM) Measure

High-Impact Strategies for Early Mathematics (HIS-EM) is a lesson-based observation tool that is designed to be used in preschool through third-grade classrooms in order to measure the quality of mathematics teaching. The aforementioned three domains (i.e., What, Who, and How) are the theoretical basis for the development of the nine HIS-EM dimensions representing teaching strategies that make a significant impact on student's mathematics learning. The HIS-EM measures the extent to which these dimensions of quality teaching practices in early mathematics, both individually and collectively, are present in an observed lesson (see Table 10).

Table 10. Description of HIS-EM Domains and Dimensions

Domain	Dimension	Definition
WHAT		Deep knowledge of foundational mathematics concepts
	Learning Objectives	Considers how well and clearly the teachers emphasizes the learning objectives , which reflect important learning and conceptual understanding, by connecting the lesson with students' prior knowledge
	Math Representations	How teacher promotes the use of multiple math representations to illustrate and connect math ideas and concepts accurately and coherently
	Concept Development	Measures the extent to which teacher's math content knowledge is accurate and coherent, Examines whether a teacher anticipates common student misconceptions, draws out key math ideas for students, and helps them generalize their understanding
WHO		Understanding of young children's typical learning pathways in mathematics and diverse students' learning needs
	Attention to Developmental Trajectories	Assesses the teacher's awareness and knowledge of the developmental trajectory for different mathematical ideas and assesses the degree to which the teachers provides feedback that promotes students' learning and clarifies student errors
	Response to Students' Individual Needs	Evaluates the teacher's awareness of and response to students' different academic needs and assesses the degree to which the teachers facilitates students' ability to actively explore and learn at their own pace by monitoring their work and adjusting the lesson as needed
	Developmentally Appropriate Learning Formats	Assesses the degree to which the instructional grouping and the pace of the lesson are appropriate and productive for the age of the students and whether the lesson is hands-on, meaningful, and connected to students' lives
HOW		Effective use of instructional support
	Planning	Considers how well the teachers selects and prepares a coherent and well-organized math lesson that helps students focus on math concepts
	Student Engagement	Assesses the degree to which the teacher facilitates opportunities for students to construct and make meaning of mathematical ideas and make use of variety of strategies to solve problems and justify their thinking
	Establishment of Math Learning Communities	Captures the degree to which the teacher's attitudes towards math and her interactions with students foster a sense of community in which all students feel welcomed to share their mathematical ideas and contribute to the lesson and classroom discourse

Finally, dimensions are explained by various observable indicators. Each dimension consists of 3 to 4 indicators of high-impact instruction. It is important to note that these indicators may not always present or equally significant in each lesson. In other words, the indicators under each dimension are not a check-list and observers evaluate dimensions holistically (see Table 11).

Table 11. Sample Domain from HIS-EM's coding guide: "What" Domain

Domain	Dimension	Indicator	Operational Definition
What: Knowledge of Foundational Mathematics Concepts	Learning Objectives	Clarity	Learning objectives are clear.
		"Big Ideas"	Learning objectives reflect conceptual understanding and important learning.
		Integrates with prior knowledge	The teacher integrates the lesson with prior knowledge.
		Reorientation statements	The teacher effectively focuses students' attention toward the purpose of the lesson.
	Mathematical Representations	Words and Gestures	Mathematical words and gestures are used frequently and correctly to illustrate concepts.
		Tools	Mathematical tools enable students to investigate concepts and represent their ideas. Connections are made between tools and mathematical concepts.
		Models	Mathematical models are accurate, varied, and help students make connections between concepts.
	Concept Development	Accuracy	The teacher displays deep, connected content knowledge.
		Anticipates common student misconceptions	The teacher anticipates common student misconceptions and successfully clarifies concepts for students.
		Deeper understanding	The lesson leads students to a deeper understanding of the concept.
		Concluding statements	The teacher concludes the lesson by summarizing mathematical concepts and helping students generalize their understanding.

Observation Procedures

This section will describe procedures related to observing a math lesson and scoring it in accordance to HIS-EM.

Observing a Math Lesson with HIS-EM. Each lesson is observed from start to finish as determined by the teacher. During the lesson, the observer watches the lesson and takes notes that are related to each dimension. These notes eventually help the observer to assign a score for each dimension. After the lesson is over, in order to complete the scoring, the observer can refer back to a manual and the notes as much as he or she needs to arrive at a score.

Scoring with the HIS-EM. HIS-EM scoring is completed immediately following the end of the lesson. The HIS-EM Observation Sheet is used by the observer to assign his scores for each dimension and write notes of justification. Observers give a score for each dimension using a 7-point scale (1 being the lowest and 7 being the highest). The dimension descriptions below provide explanations and examples of each scale at the low (1,2), middle (3,4,5) and high (6,7) ranges (see Figure 6).

Low Range		Middle Range			High Range	
1	**2**	**3**	**4**	**5**	**6**	**7**
The low range description fits the teacher and the lesson very well. All or almost all of the indicators are present and relevant to the low range.	The low range description mostly fits the teacher and the lesson, but there are also some indicators that are also in the middle range.	The middle range description mostly fits to the teacher and the lesson, but there are also some indicators that are in the low range.	The middle range description fits the teacher and the lesson well. All or almost all of the indicators are present and relevant to the middle range.	The middle range description mostly fits the teacher and the lesson, but there are also some indicators that are also on the high range.	The high range description mostly fits to the teacher and the lesson, but there are also some indicators in the middle range.	The high range description fits the teacher and the lesson well. All or almost all of the indicators are present and relevant to the high range.

Figure 6. Description HIS-EM scale.

Domain scores for each observation (i.e., What, Who, and How) are computed by averaging the appropriate dimension scores. Also, it is important to remember that the indicators under each dimension are not necessarily present, or equally significant, in each lesson. Sometimes, some of the indicators might not be relevant to a particular lesson, and the teacher should not be penalized for that. In other words, the indicators under each dimension are not a check-list and the observers need to evaluate each dimension holistically.

Training Procedures

Previously videotaped mathematics lessons have been used to create a master-coded video library for HIS-EM trainings. Next, master coding and training procedures will be explained.

HIS-EM Master Coding Procedures. Videotaped math lessons are used to code and create a master-coded video database for HIS-EM trainings come from an extensive video library that the Early Math Collaborative has built throughout the years. Coded videos met the following criteria: (a) the classroom was visible on the video, (b) audio was present, and (c) the children or the teacher (or both) were present. Videos were not coded when any one of the following conditions was met: (a) the video stopped several times, or (b) the video did not capture the overall lesson. The lessons include a variety early mathematics content (e.g., number and operations; measurement; geometry). The videographers capture what the teachers are doing throughout the lesson, but they also zoom in on students working, small group discussions, and writing on the board. Master coding involves a team of at least three expert HIS-EM coders who are part of the team created HIS-EM. Team members also have extensive training in early childhood development and education, math teaching and providing professional development to teachers. After each team member watches and scores the assigned videos, the team discusses their scores and justifications. Based on group discussion, final scores (by consensus) are assigned to each video as "master scores."

HIS-EM Training Procedures. HIS-EM coder trainees participated in a three-phase training program involving: training, reliability, and drift prevention phases. *The training phase* consists of two-day in-person training that involved: reading the HIS-EM Manual and relevant literature in early mathematics (e.g., Common Core State Standards, Big Ideas of Early Mathematics Teaching, etc.); studying the HIS-EM coding guides; and viewing, scoring, and discussing three videotaped classroom observations and HIS-EM anchor videos (a total of 18 low- and high-anchor videos) in preparation for the online reliability testing. During the last day of the training, observers practiced assigning scores for at least two videos of real mathematics lessons (grade levels tend to vary across trainings) to further develop their understanding of the HIS-EM through opportunities to ask questions and engage in discussion. *The reliability phase* involves individual coding and meetings to review and discuss trained observer's codes for a sample of eight videos. In order to complete this phase successfully, trained observers were required to meet or surpass the following criteria: (a) 80% of all observer-assigned scores had to be within one point of the master-coded scores (along the 1-7 scale) at the video level and (b) achieve reliability at least on 5 out of 8 videos. For *the drift prevention phase,* once a month during the field period, all trained and reliable observers met to re-code one videotaped mathematics lesson in order to confirm their reliability in coding. These meetings provided opportunities for observers to regularly ask questions and engage in HIS-EM-focused discussion.

Re-certification. Even experienced observers may drift from accurate use of the HIS-EM measure over time. In order to realign previously certified coders' HIS-EM observation and coding skills and to make sure they could continue to use the measure fairly and accurately, they were asked to attend to recertification training. Previously certified HIS-EM observers participate in *re-certification training* which lasted about a day and a half. In this training, observers had small group activities and discussions around HIS-EM dimensions and domains. Similar to new HIS-EM observers training, they also have a chance to code two master-coded and videotaped math lessons as a group and as an individual. *The reliability phase* for previously certified observers involved individual coding and meetings to review and discuss trained observer's codes for a sample of three videos. During this process, observers, independently, watch and code a total of 3 mathematics lessons that are previously recorded. In order to complete this phase successfully, trained observers were required to meet or surpass the following criteria: (a) 80% of all observer-assigned scores had to be within one point of the master-coded scores (along the 1-7 scale) at the video level and (b) achieve reliability at least on 2 out of 3 videos.

In response to the limitations of the currently available observation tools, HIS-EM was developed to provide a new vision for measuring quality of the early mathematics teaching practices in Pre-K through 3rd grade classrooms through observation. By building on the pedagogical content knowledge framework (PCK) by Shulman (1986), HIS-EM conceptual model introduced three domains of knowledge: (a) What—deep knowledge of foundational mathematics concepts, (b) Who—teachers' understanding of young children's typical learning pathways in mathematics and diverse students' learning needs, and (c) How—teachers' effective use of instructional support in mathematics. HIS-EM claims that these sources of knowledge (i.e., What, Who, and How) can be observed in practice and essential in determining quality mathematics teaching practices. By observing how aforementioned domains of knowledge look and sound like during mathematics instruction, HIS-EM provides a distinct approach to understand and assess the quality of mathematics teaching practices during mathematics lessons. However, having well-grounded and distinct conceptual framework might not always equally translate into strong psychometric properties. Because HIS-EM is a new and complex tool with multiple domains and dimensions, it may be vulnerable to errors across raters or errors in measurement that are systematic or constant (Olswang, Svensson, Coggings, Beilinson, & Donaldson, 2006). Therefore, its psychometric properties (i.e., reliability and validity) warrant further investigation and reporting.

Need for Further Psychometric Evidence for HIS-EM

Key indicators of the quality of any measurement instrument are its psychometric properties (i.e., reliability and validity). Validity is a prerequisite for reliability, and the relationship does not work in reverse (Gay, 1987). In other words, a scoring rubric of any instrument may cause invalid interpretations even though it is proved to be a reliable instrument. Therefore, it is important for newly developed tools to establish both their reliability and validity.

The purpose of this study is to examine the reliability and validity of the HIS-EM by using quantitative methods of analysis. Explaining the interpretation of teachers' HIS-EM score is the first step for investigating its reliability and validity. Each teacher's HIS-EM score is intended to represent the quality of his or her mathematics instruction during the mathematics lesson and is conceptualized by the degree of explicitness and clarity in instructional interactions and classroom practice occurring during mathematics lesson. Comprising such interactions were teachers' demonstrations of mathematical content and use of mathematical tools and models, opportunities for students to engage in conceptual mathematics thinking and problem-solving, teacher-provided scaffolding, usage of student errors to further understanding and teacher's efforts to build a mathematics learning community in the classroom. It involves the teacher purposefully designing learning opportunities that encourage children to explicitly think, talk, and act on real-life experiences and problems in mathematical ways. Ultimately, the quality of mathematics instruction is determined by the degree to which the teacher helps the children to meaningfully interpret foundational mathematical principles and supports the development of their fragile intuitive knowledge into the robust and transferable knowledge that marks sophisticated mathematical thinking. To support this interpretation, however, several conditions such as reliability and validity of the HIS-EM scores need to be established.

Methods for establishing HIS-EM's reliability included investigating HIS-EM's internal consistency (how well the several dimensions and domains within HIS-EM hang together) and inter-rater reliability (the extent of consensus among the HIS-EM raters). A high degree internal consistency (>.70) and high degree of consensus among the raters (>.70), would be evidence of internal consistency and inter-rater reliability respectively. Methods for establishing HIS-EM's validity included criterion-related validity (the extent of the relationship between teachers' HIS-EM scores and teachers' scores on other tool(s) measuring the similar constructs) and predictive validity (how well teachers' HIS-EM score predict their students' mathematics achievement). The Classroom Observation Scoring System (CLASS; Pianta, La Paro & Hamre, 2008), an observationally-based measure assessing the quality of teacher-student interactions and general instructional quality, was used as the criterion in this study. The CLASS was chosen as the criterion tool because it measures quality of instruction in early childhood settings; and exhibits high levels of reliability (inter-rater agreement vary between .78 and .96, and internal consistency reliabilities vary between .76 and .90) and adequate levels of criterion validity (.33 to .63) (Hamre, Mashburn, Pianta, Locasle-Crouch, 2008; Pianta et al., 2008). Furthermore, the CLASS is a well-known tool in educational field and one of the few observation tools used nationwide to assess the quality of Head Start classrooms (Hamre & Maxwell, 2011). Since both HIS-EM and CLASS taps into measuring quality of instruction, some convergence between the HIS-EM and CLASS scores is expected. Therefore, in current study, criterion-related validity of the HIS-EM was explored by examining to what extent constructs measured by HIS-EM and CLASS converge with one another. The degree of the correlations between HIS-EM scores and CLASS scores would indicate whether HIS-EM and CLASS are measuring something similar or different. Predictive validity of HIS-EM was explored by examining the relationship of the quality of early mathematics teaching measured by HIS-EM and gains in young children's mathematics achievement scores measured by Applied Problems subtest of Woodcock–Johnson Tests of Cognitive Abilities, 3rd ed., (WJ-AP; Woodcock, McGrew & Mather, 2011). WJ-III is a nationally normed battery of achievement tests that are widely used in research to assess the cognitive and academic skills of young children and as outcome measures for early childhood programs. For the purposes of this study only the Applied Problems (AP) subtest of the WJ-III will be used to measure students' achievement. The degree of the correlation between teachers' HIS-EM scores and students' learning outcomes in mathematics measured by WJ-AP will be evidence for whether HIS-EM actually measures something related to mathematics teaching and learning outcomes in mathematics.

CHAPTER 3

METHODOLOGY

This study utilizes quantitative methods to assess the criterion-related and predictive validities of the High Impact Strategies in Early Mathematics (HIS-EM) and to describe types of early math teaching the HIS-EM detects among a sample of Pre-kindergarten to 3rd teachers. Criterion-related validity of HIS-EM with the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) is explored by examining to what extent constructs measured by HIS-EM and CLASS converges and discriminates from one another. Predictive validity of HIS-EM is investigated by examining the relationship between the quality of early mathematics teaching measured by HIS-EM and gains in young children's mathematics achievement scores measured by the Applied Problems subtest of Woodcock–Johnson Tests of Cognitive Abilities, 3rd ed., (WJ-AP; Woodcock, McGrew & Mather, 2011). The profile of the quality of mathematics teaching of instructors is explored by analyzing the HIS-EM observations collected during the same time period.

More specifically, the proposed study addresses the following three research questions:

1. To what extent will constructs measured by HIS-EM and CLASS converge with or discriminate from one another?

2. Does the quality of mathematics teaching measured by HIS-EM predict children's mathematical gains?

3. What is the profile of early childhood teachers' mathematics teaching quality measured by HIS-EM?

Before explaining the proposed study further, the Innovations in Early Mathematics professional development program is introduced, with descriptions of the methods involved in recruiting participating schools, and a discussion of the characteristics of the resulting sample is discussed. Followed by this introduction, the subsequent section describes each proposed study's research design in terms of sample, instruments, procedures, and the data analysis plan.

Innovations in Early Math Project

Run by the Early Math Collaborative (The Collaborative) at Erikson Institute, the Innovations in Early Math Project was a four-year professional development (PD) program. It was designed to focus on increasing teachers' early math competencies so that they could better help their students learn. The expected outcome was that students from pre-kindergarten to 3rd grade would meet or exceed the state learning standards in mathematics.

Teacher and Student Recruitment

Participants were recruited from 16 public schools in a large Midwestern city in the U.S. Assignment to treatment versus comparison conditions occurred at the school level, resulting in 8 treatment schools and 8 comparison schools. Key personnel of the

Collaborative formed an implementation team and conducted all recruitment efforts to select both treatment and comparison schools. The recruitment process involved a number of activities, including identifying eligible schools in need of instructional support in mathematics, contacting these schools' administrators and teachers, recruiting an adequate sample based on the project's goals and design, and retaining the participants until study completion.

A yearlong effort to recruit participant schools started by contacting network[2] school district leaders to inform them about the Innovations Project's research aims and activities. Interested district leaders then earmarked potential schools for participation that contained a significant student population with high needs for mathematical instructional support. The implementation team contacted the administrators of the recommended schools to determine their interest in this project. If the school's administrator expressed such interest in participating, an on-site visit was made to gather more information about the school, its atmosphere for collaborative work, and its student population (i.e., students' mobility rate). If more information was required, another school visit was scheduled. In addition to administrative interest, the implementation team also gathered information from teachers at recommended schools about their interest in participating in the study. Therefore, during school visits, the team met with teachers and informed them about the program and what it entailed. None of the administrators at any of the schools were involved or participated in these meetings with

[2] District-run schools are organized into networks, which provide administrative support, strategic direction, and leadership development to the schools within the network. There are a total of 13 networks that manage schools in various different geographic regions of Chicago.

teachers. All teachers from pre-kindergarten through 3rd grade were contacted and encouraged to participate. Teachers' interest in this project was gathered via a survey asking them to respond to a sliding scale of "strong unwillingness to participate" to "strong willingness to participate" anonymously. The implementation team chose 8 treatment schools based on teachers willingness, administrative support, and school level characteristics (e.g., student mobility). The selection process took approximately 6 months between January and June 2011.

After treatment schools were recruited, the implementation team began recruiting comparison schools. A total of 65 non-treatment schools were identified, from which the 8 comparison schools could be chosen. Decisions concerning which schools should be selected as comparison schools were made based on the degree to which they were good matches for the already selected treatment schools. For each treatment school, a matched comparison school was chosen from within the same network by following similar procedures (i.e., contacting district leaders, meeting with school administrators and teachers) and using propensity score matching techniques to ensure treatment and comparison schools were statistically comparable (Stuart, 2010). The estimated propensity model was developed using school-level variables including: a) the percentage of 3rd grade students who met math standards in 2009, b) the percentage of 3rd grade students who exceeded math standards in 2009, c) the percentage of students who were English Language Learners, d) the percentage of students who were identified as minority, e) the percentage of students receiving free or reduced-price lunch, and f) students′ mobility rates. Schools with acceptable propensity scores that showed willingness to participate at both administrator and teacher levels were recruited as comparison group schools over the course of 3 months (June, 2011 to August, 2011).

Two out of the 8 selected comparison schools dropped out after selection prior to pretest data collection. Thus, 2 replacement comparison schools were then selected from the pool of 57 non-treatment schools based on their school-level baseline variables. The resulting sample was comprised of a total of 16 schools from 6 networks. Of the 16 schools, 8 were treatment schools and 8 were comparison schools with comparable school-level baseline variables (see Table 12).

Table 12. Descriptive Statistics for School-level Matching Characteristics

Baseline School – Level variables	Overall *M* (*SD*)	Treatment *M* (*SD*)	Comparison *M* (*SD*)	ES	Significance
% free-reduced price lunch	.92 (.06)	.92(.04)	.91 (.07)	.26	.62
% meet state math standards	.49 (.09)	.50 (.08)	.48 (.11)	.17	.74
% exceed state math standards	.23 (.15)	.21 (.12)	.24 (.17)	-.21	.68
% English language learners	.25 (.15)	.19 (.10)	.33 (.11)	-1.37	.02
Mobility rate	.17 (.05)	.17 (.07)	.17 (.02)	0	1
% Hispanic	.59 (.30)	.62 (.33)	.55 (.28)	.22	.67
% Black	.26 (.34)	.29 (.37)	.23 (.32)	-.17	.74

As shown in Table 12, about 92% of the students in participating schools received free-reduced price lunch. In both treatment and comparison group schools, about half of the students were not meeting the state standards in mathematics on the Illinois State

Achievement Test (ISAT)[3]. Generally, about one student in every five students at each participating school exceeded the state standards. On average, one-fourth of the student populations in participating schools were English language learners. The students' mobility rate was 17%. On average, Hispanic and Black students represented about 59% and 26% of the whole population respectively. Statistics presented in Table 12 suggest that the resulting sample consists of early childhood teachers working with students from low-income families who require more mathematical support.

Sample Description[4]

Teacher Participants. A total 210 teachers participated in the larger study during Year1. The number teachers from each school ranged from 6 to 18 (*M*=13). The distributions were comparable across grade levels and between comparison and interventions schools. There were about 114 teachers from each elementary grade and about 85 teachers from Pre-K and Kindergarten respectively, and 11 teachers (5% of the sample) were working in mixed age classrooms (e.g., Kindergarten and 1st grade split) (see Table 13).

Table 13. The Distribution of Teachers by Grade Level

Grade	Pre-K	K	K -1 Split	1st	1 - 2 Split	2nd	2 - 3 Split	3rd
Overall	39	46	3	42	6	34	2	38
%	18.6	21.9	1.4	20.0	2.9	16.2	1.0	18.1

Note: *N*=210
This is corresponding to the teachers who were observed with HIS-EM.

[3] The Illinois State Achievement Test (ISAT) measures individual student achievement relative to the Illinois Learning Standards.

[4] This sample description only included the relevant data from the year 1 of the larger study.

As shown in Table 14, 95.7 % of the participants were females and the majority (60%) was between 25 and 44 years old. More than one quarter of the sample had been teaching for less than five years. About half of the sample identified themselves as Caucasian/White, one third of the sample was Hispanic/Latino, and one tenth was Black. All teachers were certified to teach and had varying degrees of professional development experiences in early mathematics.

Table 14. The Background Information of Participating Teachers

Age Span	24 and under	25-34	35-44	45-54	55-64	65 and over	
%	2.7	30.4	29.5	25.0	9.8	2.7	
Gender	Female	Male					
%	95.7	4.3					
Ethnicity	African-American or Black	American Indian or Alaska Native	Asian	Caucasian or White	Hispanic or Latino	Native Hawaiian or other Pacific Islander	Other
%	10.6	.9	7.1	46.0	31.9	-	1.8
Years of Teaching	Mean	SD	Range	Less than 5	6 to10	11 to 15	15 and above
%	12.85	9.27	1 to 41	27.5	20.9	19.8	31.9
Certificate	Early Childhood Teacher Certificate (Type 04)	Elementary Education Certificate (Type 03)	Bilingual Endorsement	Special Education Certificate	Others		
%	32.8	69.9	39.9	16.9	71.0		
Early Math PD hours teachers attended	Mean (SD)	SD	Range	0	1 to 5	6 to 15	More than 15
%	8.87	12.95	0 to 80	29.7	25.3	28.6	16.5

Note= *N*=115~182[5]

Student Participants. An estimated 6,000 children were enrolled in PK-3 across the 16 participating schools in 2011–12. Of those, 2,609 (43%) children were consented. Of the 2,609 children whose parents consented for them to participate in the study, EMC research team attempted to assess between 7 and 10 children per classroom teacher

[5] Demographic information for all participating teachers was not available and varied from teacher to teacher.

resulting sample of 1,551 children assessed at pretest across 188 teachers. 9% of the children who were assessed in pre-test were not able to be assessed at post-test due to various reasons (e.g., family moved, child absences). Of the 1,551 children who had pretest assessments, 1,404 were assessed at post-test (or 91 % of the original pretest sample). Table 15 shows the overall child-level data such as, gender, age in months, and grade level (reported by parents and children′s school) at both pre-test and post-test.

Table 15. The Distribution of Students by Gender, Grade Level, and Age at Pre-test

	n	%
Gender		
Boys	696	49.6
Girls	708	50.4
Grade level		
Pre-K	285	20.3
K	340	24.2
1st	295	21.0
2nd	225	16.0
3rd	259	18.4
Age at pretest (in months)		
40 to 48	21	1.5
49 to 60	217	15.4
61 to 72	326	23.2
73 to 84	306	21.7
85 to 96	239	17.0
97 to 108	233	16.5
109 to 120	42	2.9
120 to 128	2	0.14

Note: *N*=1,404

Intervention

The Early Math Innovations Project focused on improving students' mathematics achievement by increasing teachers' foundational mathematics knowledge, changing their attitudes and beliefs about teaching mathematics, and improving the quality of their mathematics teaching. During the project, treatment schools received PD training in early mathematics that consisted of four components: (1) learning labs—interactive learning sessions that helping teachers gain understanding of foundational content and related instructional strategies in mathematics; (2) on-site coaching—individualized planning, observing, and reflecting guided by skilled and math-knowledgeable coaches who helped teachers move classroom practice to new levels of pedagogy; (3) school-based learning groups—a venue for collaboration among teachers from the same grade as well as across grades to examine students' work, share effective instructional strategies, and study the Illinois Learning Standards for Mathematics (ILSM) and performance descriptors outlined by Common Core Standards State Standards in Mathematics (CCSS-M) and integrate them into their math curriculum; and (4) leadership academies—designed to increase administrators' awareness of what constitutes effective and high quality mathematics teaching (see Figure 7). Comparison schools were not provided with any PD by the Collaborative during the study period.

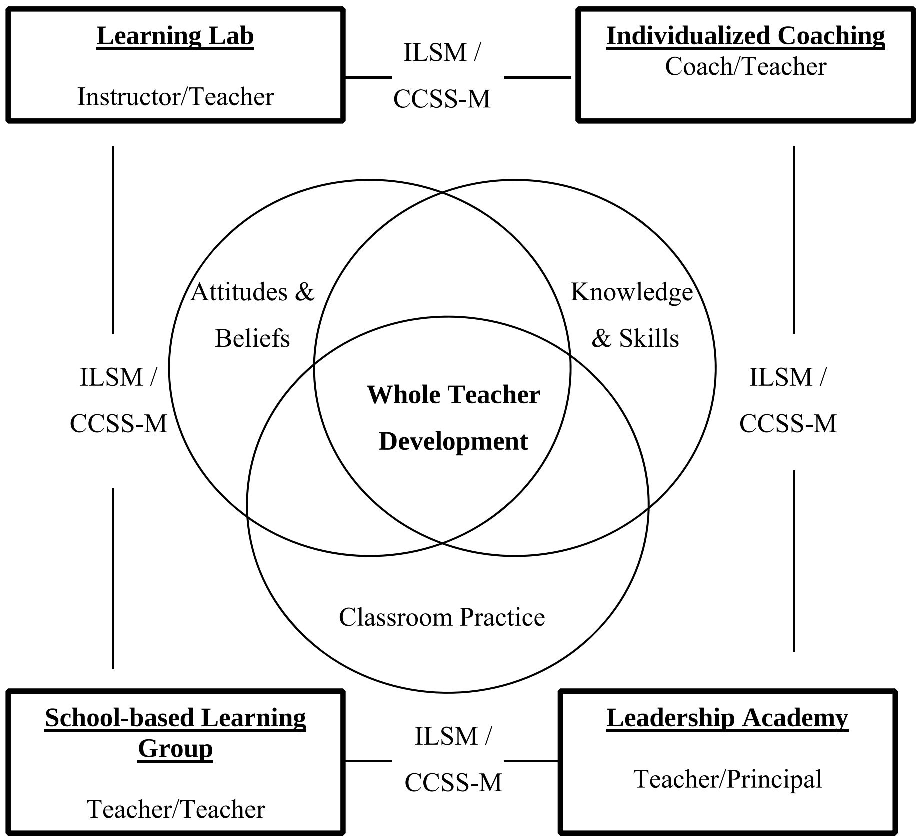

Figure 7. Program Components and Conceptual Framework for Early Math Innovations Project.

Data Collection

Over the course of the project, data was collected at both teacher and student levels at predetermined time points (i.e., fall 2011, spring 2012, spring 2013, spring 2014, and spring 2015). As part of the Early Math Innovations Project, all the participating teachers (both treatment and comparison) were asked to schedule a math lesson observation (i.e., HIS-EM) and complete an online survey (i.e., PCK-EM,[6] ABC-EM,[7] &

[6] Teachers' pedagogical content knowledge is measured via online administration Pedagogical Content Knowledge in Early Mathematics survey (PCK-EM).

About My Teaching)[8] at each data collection point during the course of four years. Consenting students in participating teachers' classrooms were assessed at the first three time points (i.e., fall 2011, spring 2012, and spring 2013). No student level data (i.e., WJ-AP[9] & TEAM)[10] was collected during the rest of the study (see Table 16).

Table 16. Data Collection Timeline for Teacher and Child Measures

	Teacher or Child Data	Pre (Fall 2011)	Post (Spring 2012)	Follow-Up (Spring 2013)	Follow-Up 2 (Spring 2014)	Follow-Up 3 (Spring 2015)
HIS-EM	Teacher	√	√	√	√	√
PCK-EM	Teacher	√	√	√	√	√
ABC-EM	Teacher	√	√	√	√	√
About My Teaching	Teacher	√	√	√	-	-
WJ-AP & TEAM	Child	√	√	√	-	-

Compensation

Regarding the research incentives, all teachers (i.e., both treatment and comparison group teachers) who completed on-line surveys (e.g., About My Teaching) received a $50 gift card after submitting their responses at each time point. Participating teachers at the treatment schools also received a stipend for attending learning labs

[7] Teachers' attitudes and beliefs about teaching mathematics are measured via online administration of Attitudes, Beliefs and Competence in Early Mathematics survey (ABC-EM).

[8] Teachers' background in teaching and demographic information is gathered via online administration of About My Teaching survey.

[9] Woodcock Johnson-III Applied Problem (WJ-AP; subtest #10), indicating students' performance in mathematics.

[10] Tools for Early Assessment in Math (TEAM; Clements, Sarama, & Wolfe, 2011) measures core mathematical abilities of young children across 19 learning trajectories.

outside their work schedule and were given children's story books. Comparison school teachers, on the other hand, received different types of incentives. At the classroom level, they received two science-themed books per year. In addition, comparison teachers were also rewarded with membership to either "RAZ Kids" or "Reading A-Z," (an online independent reading site that offers several interactive, leveled and printable e-books written in multiple languages). End of study awards for participating comparison schools included a set of 36 children's books that support the teaching and understanding of foundational mathematics. Participating teachers at the comparison schools were also provided with a total of 4 professional development vouchers during the first 3 years; and 8 during the final year of the project. These vouchers were issued by the Erikson Institute the Collaborative was affiliated with and could be used to participate in any of the non-math-related PDs offered by the Erikson Institute itself. Once the study was completed, the Collaborative team also provided two half-day workshops to all pre-kindergarten to 3rd grade teachers at participating comparison schools.

Research Design for the Present Study

The present study only utilized the data collected in year 1 of the Collaborative's study (i.e., pre-test in fall 2011 and post-test in spring 2012). This set of data provided the possibility to test research questions regarding: (1) HIS-EM's criterion-related validity, (2) HIS-EM's predictive validity, and (3) HIS-EM profiles among observed teachers (see Table 17). Below, the research design for answering each research question is described.

Table 17. Research Design of the Three Research Questions

Research Question	Teacher Sample	Teacher Measure	Student Sample	Student Measure	Data	Data Collection Point
Criterion-related validity of the HIS-EM with the CLASS	Subgroup of Treatment	HIS-EM, CLASS, About My Teaching	NA	NA	Video-taped math lessons	During the first year of the study
Predictive validity of HIS-EM	Comparison	HIS-EM, About My Teaching	Comparison	WJ-AP	Live in-class observations and one-to-one student assessment	Pre-test for teachers, post for students (controlling for students' pretest scores)
Profile of HIS-EM	Treatment and Comparison	HIS-EM	NA	NA	Live in-class observations	Pre-test

Note:
NA: Not applicable.
Treatment: Teachers participating in the treatment condition, who received on and off site coaching support in early mathematics from the Collaborative.
Comparison: Teachers participating in the comparison condition, who did not receive any kind of on and off site coaching support in early mathematics from the Collaborative.
HIS-EM: High Impact Strategies in Early Mathematics observation tool, indicating the quality of teaching mathematics.
CLASS: Classroom Assessment Scoring System, indicating the global quality of teaching.
About My Teaching: Teachers' background in teaching and demographic information is gathered via online administration of About My Teaching survey.
WJ-AP: Woodcock Johnson-III Applied Problem (subtest #10), indicating students' performance in mathematics.

As outlined by Table 17, the current study addresses three major research questions. Each research question uses a different methodology and draws from different samples. Therefore, the following section is presented according to each research question with subsections for sample description, instruments, procedures and analysis.

Research Question 1: Establishing Criterion-related Validity of HIS-EM with CLASS

This research question investigates the criterion-related validity of the HIS-EM with the Classroom Assessment Scoring System (CLASS; Pianta, et. al., 2008) by examining to what extent constructs measured by HIS-EM and CLASS converge with or discriminate with one another.

Sample Description

The goal of this study was to examine the relationship between two different observation tools, not the effects of intervention. Therefore, teachers in the comparison condition were excluded from the sample. Teachers participating in the treatment condition, who received on site coaching support from the Collaborative and were videotaped by the Collaborative's research team during the first year of the larger study, were eligible to participate in this study. Of the 108 teachers in treatment group, 91 of them were videotaped by trained videographers at least once while teaching mathematics during the first year of the larger study.

To determine which videos to code, captured videos were further analyzed based on HIS-EM video coding guidelines. Initial screening of the videos based on these coding guidelines revealed that most of the videos gathered from preschool and kindergarten classrooms met the coding requirements set by the HIS-EM. Videos gathered from higher grades did not meet the requirement set by coding guideline (e.g., does not capture the lesson from beginning). More specifically, of the 41 preschool and kindergarten teachers in the treatment group, a total of 27 teachers met the requirements set by the HIS-EM tool to participate and composed the overall sample for this study (see Figure 8).

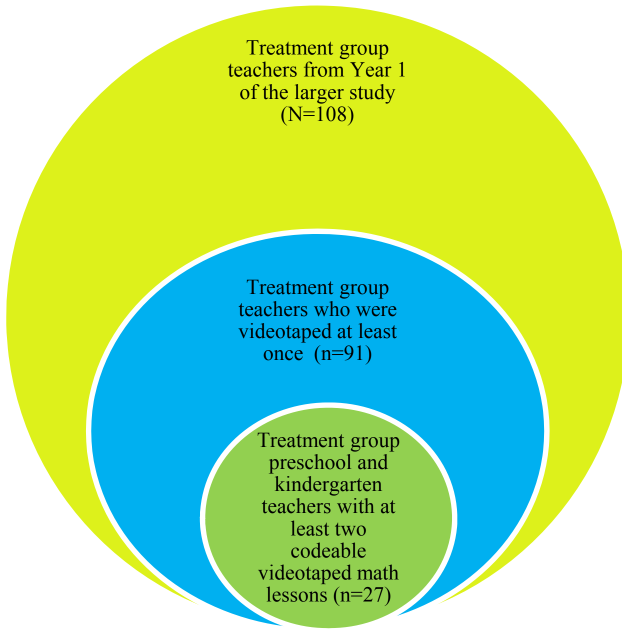

Figure 8. Description of the sample for examining criterion-related validity of HIS-EM.

Of the 27 teachers (100% female) included in the final sample, only 20 of them have available demographic information. Analysis of available demographic information (*n*=20) revealed that all of the teachers had a bachelor's degree and had a mean of 13.1 years (*SD* = 8.63, range = 1-30) of experience working professionally as early childhood educators at the time of the study.

Instruments

Two observation tools were used to measure the quality of instruction: High Impact Strategies in Early Mathematics (HIS-EM; Early Math Collaborative, 2011) and Classroom Assessment Scoring System (CLASS; Pianta, et. al., 2008).

Classroom Assessment Scoring System (CLASS) is a system for observing and assessing the quality of interactions between students and teachers in a classroom (Pianta, LaParo & Hamre, 2008). The CLASS has a total of 10 dimensions, each of which is scored on a 1–7 scale: low (1–2), medium (3–5), and high (6–7). Anchor point descriptions for each dimension guide raters in selecting an appropriate score level. The ten dimensions are used to measure quality in three domains: (1) Emotional Support, (2) Classroom Organization, and (3) Instructional Support. Each dimension only contributes to scores on one domain.

The emotional support domain is assessed through scoring the nature of the climate (positive or negative), the sensitivity of the teacher, and the regard the teacher holds for various student perspectives, ideas, interests, and skills. Assessing behavior management, productivity, and instructional format on the measure provides insights into the Classroom Organization domain. The "Instructional Support" domain is ascertained by looking at how concepts are developed, feedback is provided, and language is modeled (see Table 15).

Adequate criterion validity has been demonstrated for the CLASS (.33 to .63), including associations with other measures of classroom quality such as ECERS-R (Hamre, Mashburn, Pianta, Locasle-Crouch, 2008; Pianta et al., 2008). Inter-rater agreement has been reported to vary between .78 and .96, and internal consistency reliabilities varied between .76 - .90. Factor analysis studies revealed mixed results. While some support the three domain structure of the measure (Malmberg, Hagger, Burn, Mutton, & Colls, 2010), others suggested that either a 3 factor or single factor model at the teacher level are both plausible models (McCaffrey, Yuan, Savitsky, Lockwood, & Edelsen, 2014).

Table 18. Description of CLASS Domains and Dimensions

Domain	Dimension	Definition
Emotional Support		Support for social-emotional functioning in classroom
	Positive Climate	Emotional connection, relationships, and positive communications among teachers and children
	Negative Climate	Level of negativity in the interactions among teacher and children in the classroom
	Teacher Sensitivity	Teacher responsiveness to children's academic, social, emotional, and developmental needs
	Regard for Student Perspectives	Teacher-child interactions and classroom activities that emphasize children's interests and ideas
Classroom Management		Behavior management and classroom organization
	Behavior Management	Teacher use of effective methods to prevent and redirect children's misbehavior
	Productivity	Teacher management of time to maximize children's learning opportunities
	Instructional Learning Formats	Teacher facilitation of children's engagement through interesting activities, instruction, and materials
Instructional Support		Promoting language skills and cognitive development
	Concept Development	Teacher use of instructional activities that promote children higher order thinking skills
	Quality of Feedback	Teacher use of feedback focused on expanding children's learning and understanding
	Language Modeling	Teacher use of language-stimulation and language-facilitation techniques while interacting with children

Note. Definitions and examples were based on the CLASS Pre-K manual (Pianta et al., 2008).

Procedure for Data Collection

For this study, previously videotaped math lessons for coaching purposes were coded by using HIS-EM and CLASS. All data was collected between fall 2011 and spring 2012. The description below further explains how existing data was collected.

Video-taped Math Lessons. In order to examine the convergent or discriminant patterns between HIS-EM and CLASS, math lesson videos of a sub-sample of treatment group pre-kindergarten and kindergarten teachers from Year 1 of the larger study were coded using HIS-EM and CLASS measures (n=27). Of the 130 teaching videos available, a total of 54 from the 27 teachers in the subsample were coded for this study (i.e., two videos per teacher). On average, identified videos were 23.81 minutes long (SD = 10.06, range = 9-62). The videotaped lessons included a variety of early mathematics content (e.g., number and operations; measurement; geometry). The videographers captured what the teachers were doing throughout the lesson, but they also zoomed in on students working, small group discussions, and writing on the board.

The 54 videotapes were coded by a group of 3 trained HIS-EM observers who hold a current HIS-EM certification at the time of this study and a separate group of 3 CLASS observers who hold a valid CLASS certification at the time of this study. Observers were instructed to observe, take notes, and score the videos independently and to not discuss the observation/scoring until the scores have been submitted and the score sheets have been collected.

Data Analyses

Analysis for Study 1 included descriptive summaries of the frequencies and rates of codes observed in the 54 video-recorded math lessons, as well as bivariate correlations for tests of association. Also, linear and multiple regression analyses were run to test the direction of the relationships between HIS-EM and CLASS domain scores.

Research Question 2: Establishing Predictive Validity of HIS-EM

This research question attempts to address whether the quality of mathematics teaching measured by HIS-EM predicts children's mathematical gains over the course of an academic year.

Sample Description

Teacher Participants. 16 public schools (8 treatment and 8 comparison schools) from 6 networks in a large Midwestern city in the U.S. participated in the Innovations study. It is hypothesized that teachers' instructional quality as measured by HIS-EM might change as they participate in the intervention and further impact their students' gains in mathematics. To minimize the effects of intervention on the relationship between the quality of math instruction measured by HIS-EM and students' gains, teachers in the treatment condition in year 1 of the larger study will be excluded from this sample. Therefore, the analytic sample for this study will only involve 92 teachers in the comparison condition who were part of the larger study between the 2011 and 2012 academic years (i.e., Year 1 of the larger study). Even though excluding treatment group teachers may result in more robust research design, it also introduces a new challenge by decreasing the available sample size significantly.

Because the limitations created by a small sample size can have profound effects on the outcome and its statistical power, teacher data available from all grade levels (i.e., Pre-K to 3rd grade) were included in this study. Descriptive analysis of the available data about the sample revealed that the number of teachers from each comparison school ranged from 3 to 12. There are 37 teachers from the primary grades (e.g., first, second, and third) and 36 teachers from pre-kindergartens and kindergarten (see Table 19).

Table 19. Distribution of Teachers by Grade Level

	Pre-K	K	1st	2nd	3rd	1-2	Total
Comparison	15	21	11	12	13	1	73
%	20.5	28.8	15.1	16.4	17.8	1.4	100

Note: N= 94
This is corresponding to the teachers who were observed with HIS-EM.
Pre-K: Pre-kindergarten.
K: Kindergarten.
1-2: 1st and 2nd graders mixed class.
1st: First grade.
2nd: Second grade.
3rd: Third grade.

Student Participants. Assignment of condition for students also occurred at the school level such that all children in the treatment schools with participating teachers received the treatment and all children in the comparison schools did not receive any treatment. Participating teachers in each research condition were provided with consent forms prepared by the Collaborative’s research team. Later, teachers were asked to distribute these forms to their students’ parents and ask them to sign if they would be interested in allowing their children to participate in the study. Consenting students were included in this study, if they were also; a) enrolled in the classroom of the participating teacher, b) able to complete the student assessments in English or Spanish, c) 4-years-old or older by the time they were first assessed in fall 2011. Children were excluded from the study; a) if their parents did not consent, b) if they were unable to complete the assessment in English or Spanish; c) if they had an Individualized Education Plan (IEP) and/or a 504 Plan[11], and/or d) if they were not present on the days the team visited the school at pretest. Of the 2,609 children whose parents consented for them to participate in the study, a sample of 1,404 was assessed at both pre-test and post-test. Of 1,404 students, 546 of them were in the comparison group and data gathered from this group of students will be used and analyzed in this study. Table 20 shows the overall child-level and comparison group child-level data in terms of their grade level.

Table 20. Distribution of Students by Grade Level

	Pre-K	K	1st	2nd	3rd	Total
Comparison	131	160	89	75	91	546
%	24	29.3	16.3	13.7	16.7	100

Instruments

Assessment tools that were used in this study include HIS-EM (Early Math Collaborative, 2011), WJ-AP subtest (Woodcock, McGrew, & Mather, 2011) and "About my Teaching" on-line survey.

The Woodcock-Johnson-III Applied Problems subtest (WJ-AP) is an individually administered norm-referenced test that measures skills in analyzing and solving practical math problems with a total of 60 items. It is the 10th subtest of

[11] A 504 Plan makes sure a child with special health care needs has the same access to education as other children. It is supported by the federal civil rights law, Section 504 of the Rehabilitation Act of 1973.

Woodcock Johnson-III (Woodcock, et al., 2011). The test administered verbally presents items involving counting, telling time or temperature, and problem-solving. Items are ordered in terms of their age-appropriateness. Testing begins with an item corresponding to the subject's age and is discontinued after 6 consecutive errors. The score is determined by summing the number of correct responses. Internal alpha reliability estimates are reported as .88 to .94 for English speaking children ages 4 to 7 years.

About my teaching is an online survey collecting participating teachers' demographic information and teaching and learning experiences in regard to early mathematics education. The questions included in the survey were aimed to elicit information about a participating teacher's educational background, experience in participating pre-service and in-service workshops teaching mathematics, as well as his or her experiences working with English Language Learners (ELL). For example, teachers were asked to answers questions such as: *How many years have you been teaching?; About how many hours of in-service math education workshops have you taken in the last two years?; and How many years of experience do you have working with ELL students in a classroom setting?*

Procedure for Data Collection

All teacher observations and child assessments were collected between fall 2011 and spring 2012. The description below further explains how existing data was collected.

HIS-EM Classroom Observations. Trained observers conducted live in-class observations in the fall 2011 (pre-test) and spring 2012 (post-test) at each participating school. One observation per classroom at each time point was planned. Each observation at each time point is considered a snapshot representing how mathematics instruction may function across a given school year. The program coordinator of the Collaborative scheduled all the observations in coordination with the participating teachers. All classroom observations were scheduled in advance and conducted during the time the teacher allocated to teach mathematics or the mathematics lesson time period. Scheduled observations were not specific to mathematical content (e.g., number and operations or geometry or etc.), or a particular instructional day (e.g., start or end of a weekly math unit). Observers remained in each classroom for the duration of the mathematics lesson.

Applied Problems subtest of Woodcock–Johnson Tests of Cognitive Abilities, 3rd ed., (WJ-AP). Young children's mathematical achievement was assessed via WJ-AP subtest (Woodcock, McGrew, & Mather, 2011) in the fall 2011 (pre-test) and spring 2012 (post-test) in each participating classroom. Consenting teachers were asked to distribute consent forms to parents of their students in the class. When necessary, consent forms were translated into languages other than English, such as Spanish, Polish, Urdu and Arabic. Because only the children whose parents consented to the study could be assessed, the number of students assessed in each classroom was not consistent. However, the total number of children from each classroom never exceeded 10. For example, if more than 10 students gave consent in any given classroom, only 10 students among all the consenting children were randomly selected and assessed. If the number was not more than 10, then all the consenting children were assessed.

Data Analyses

In order to examine the relationship between quality of mathematics teaching measured by HIS-EM and students' learning gains in mathematics over a school year, three-level hierarchical linear modeling (HLM) analyses (Raudenbush & Byrk, 2002) was conducted by using the HLM program. Hierarchical Linear Modeling (HLM) is a type of regression model often used for analyzing education data sets because they tend to include multiple layers of data that are correlated with one another because they share similar traits (e.g., students from the same classroom and schools are similar in their traits) (Raudenbush & Byrk, 2002). In this analysis, students (Level 1) were nested within teachers (Level 2), who were nested within schools (Level 3). Using three-level HLM, relationships between students' math achievement and quality of mathematics teaching was estimated after controlling for school variations.

Research Question 3: The HIS-EM Profile of Early Childhood Teachers

This research question examines what kind of HIS-EM teaching profiles exist among early childhood teachers.

Sample Description

16 public schools (8 treatment and 8 comparison schools) from 6 networks in a large Midwestern city in the U.S. participated in the Innovations study. The analytic sample for this study will involve a total of 210 teachers who were part of the study in between 2011 and 2012 academic year (i.e., Year 1 of the larger study), regardless of their research condition. Initial analysis of the available pre-test data about the sample revealed that the number of teachers from each grade ranged from 2 to 46. There are 85

teachers from pre-kindergarten and kindergarten while the rest are from primary grades (see Table 21).

Table 21. Distribution of Teachers at Pretest by Grade Level

	Pre-K	K	1st	2nd	3rd	K-1	1-2	2-3	Total
Overall	39	46	42	34	38	3	6	2	210
%	18.6	21.9	20.0	16.2	18.1	1.4	2.9	1.0	100

Note: *N*=210
This is corresponding to the teachers who were observed with HIS-EM.
Pre-K: Pre-kindergarten grade.
K: Kindergarten grade.
K-1: Kindergarteners and 1st graders mixed class.
1-2: 1st and 2nd graders mixed class.
2-3: 2nd and 3rd graders mixed class.
1st: First grade.
2nd: Second grade.
3rd: Third grade.

Instruments

HIS-EM (Early Math Collaborative, 2011) will be used as an assessment tool to determine the existing math teaching profiles among observed early childhood teachers.

Procedure for Data Collection

The same data collection procedures applied in previous section (i.e., research question 2) for teacher level data used.

Data Analyses

The 210 math lessons were observed and coded by a group of nine trained HIS-EM observers who hold a current HIS-EM certification at the time of this study (fall 2011). Observers were instructed to observe, take notes, and score the lessons immediately after the observed lesson was concluded. Descriptive analyses were conducted to report mean, standard deviation, and range of each HIS-EM dimension and domain by running descriptive analyses on SPSS. The extents to which HIS-EM domains tend to change together were calculated to examine the relationships between HIS-EM domains. To investigate the profile of the quality of mathematics teaching in early childhood classrooms measured by HIS-EM, a *two-step cluster analysis* was run using SPSS. This analysis was enable investigation and identification of groups of teachers whose teaching showed similar levels of mathematics teaching measured by HIS-EM.

CHAPTER 4

RESULT

This chapter presents the results of statistical analyses concerning the data that address the study's three research questions: (1) criterion-related validity of HIS-EM with CLASS, (2) predictive validity of HIS-EM (whether teachers' HIS-EM scores predict students' mathematics outcomes), and (3) what kind of HIS-EM profiles exist among early childhood teachers.

Setting the Stage for Further Investigation: Inter-rater Reliability of HIS-EM and CLASS

When using observational instruments which involve judgments or ratings by observers, a reliable measure will require consistency between the coders. To demonstrate consistency among observational ratings provided by multiple coders in this study, 14 videos of the 54 videos (26%) were randomly assigned to a pair of HIS-EM and CLASS observers and were double coded (twice for HIS-EM and twice for CLASS). Inter-observer agreement levels between the coders for this set of videos (n=14) were analyzed and reported at two levels; percent adjacent agreement and intra-class correlations.

Percentage of adjacent agreement. It was calculated to provide information regarding the precise agreement between coders. The percentage of scores across the 14 videos for which two observers were within 1 point of each other on the 7-point scale (i.e., percent adjacent agreement) was calculated for each HIS-EM and CLASS dimension. For HIS-EM, the percent adjacent agreement (equivalent or within one point) across all HIS-EM dimensions was 80.1% (met the reliability requirements set by HIS-EM; requires at least 80%) and by dimension ranged from 71.4% to 92.8%. For CLASS, the rate of adjacent agreement across all CLASS dimensions was 85% and met the reliability requirements set by the CLASS (requires at least 80%). The rates of adjacent agreement by dimensions ranged from 64.3% to 100% (see Table 22).

Table 22. The Rates of Adjacent Agreement for HIS-EM and CLASS Dimensions

Measure		Percent Adjacent Agreement (%)
HIS-EM	Learning Objectives	78.5
	Math Representations	78.5
	Concept Development	78.5
	Attention to Developmental Trajectories	85.7
	Response to Students' Individual Needs	71.4
	Developmentally Appropriate Learning Formats	78.5
	Planning	92.8
	Student Engagement	78.5
	Establishment of Math Learning Communities	78.5
	HIS-EM (Overall)	80.1
CLASS	Positive Climate	100
	Negative Climate	100
	Teacher Sensitivity	92.8
	Regard for Student Perspectives	100
	Behavior Management	64.3
	Productivity	100
	Instructional Learning Formats	100
	Concept Development	100
	Quality of Feedback	92.8
	Language Modeling	100
	CLASS (Overall)	85

Note: N=14

Intra-class correlation (ICC). Analyses of intra-class correlations between variables provide an estimate of the relationship between two variables of the same unit or construct, the index is commonly used as a measure of inter-rater reliability (Field, 2009). Two-way random effects model of intra-class correlations (ICC) analyses, in which teachers and raters are treated as random, were conducted to determine inter-rater reliability of the quality assessment scores for HIS-EM and CLASS dimensions. The results of the ICC analyses revealed that intra-class correlations for HIS-EM ranged from .56 to .87 while those for CLASS ranged from .31 to .83. Based on commonly cited cutoff points by Cicchetti (1994)[12], most of the ICCs for HIS-EM dimensions reached either "good" or "excellent" reliability except for "Student Engagement" dimension which was fair. This indicates that HIS-EM coders had a high degree of agreement and suggests that HIS-EM dimensions were rated similarly across coders (see Table 23).

[12]Cicchetti (1994) suggests commonly-cited cutoffs for qualitative ratings of agreement based on ICC values, with poor for values less than .40, fair for values between .40 and .59, good for values between .60 and .74, and excellent for values between .75 and 1.0.

Table 23. Intra-class Correlations for HIS-EM Dimensions

Measure	Dimension	α	ICC
HIS-EM	Learning Objectives	.787	.648
	Math Representations	.870	.770
	Concept Development	.844	.729
	Attention to Developmental Trajectories	.933	.874
	Response to Students' Individual Needs	.812	.684
	Developmentally Appropriate Learning Formats	.762	.616
	Planning	.903	.824
	Student Engagement	.716	.558
	Establishment of Math Learning Communities	.831	.711

Note: N=14

Most of the ICCs for CLASS dimensions showed variations and ranged from "fair" to "excellent" reliability except for the "Behavior Management," dimension. ICC estimates for this dimension was "poor". These results indicate that CLASS coders had a high degree of agreement in general and suggest that CLASS dimensions were rated similarly across coders, except for "Behavior Management" dimension (see Table 24).

Table 24. Intra-class Correlations for CLASS Dimensions

Measure	Dimension	α	ICC
CLASS	Positive Climate	.729	.574
	Negative Climate	.787	.649
	Teacher Sensitivity	.797	.662
	Regard for Student Perspectives	.899	.817
	Behavior Management	.479	.315
	Productivity	.691	.528
	Instructional Learning Formats	.910	.835
	Concept Development	.845	.732
	Quality of Feedback	.746	.595
	Language Modeling	.925	.860

Note: N=14

Establishing Criterion-Related Validity of HIS-EM with CLASS

Evidence of criterion-related validity of the HIS-EM was explored by examining the relationship between HIS-EM, a subject-specific measure of instructional quality in early mathematics teaching, and the Classroom Assessment Scoring System (CLASS), a global measure of instructional quality. Sample for this study consisted of math lesson videos of a sub-sample of pre-kindergarten and kindergarten teachers in treatment group from Year 1 of the larger study (*n*=27). Of the 130 teaching videos available, a total of 54 from the 27 teachers in the subsample were coded for this study (i.e., two videos per teacher). On average, identified videos were 23.81 minutes long (*SD* = 10.06, range = 9-62). The videotaped lessons included a variety of early mathematics content (e.g., number and operations; measurement; geometry). The videographers captured what the teachers were doing throughout the lesson, but they also zoomed in on students working, small group discussions, and writing on the board. In order to investigate criterion-related validity of the HIS-EM with CLASS, several statistical analyses were performed including; determining the internal consistency of HIS-EM and CLASS dimensions, determining the distribution of HIS-EM and CLASS scores across the sample, and exploring the relationship between HIS-EM and CLASS and its domains.

Internal Consistency

The degree of the consistency of HIS-EM dimensions with one another and CLASS dimensions with one another were analyzed by running basic descriptive analyses and scale reliability analyses such as Cronbach's alpha. Cronbach (1951) defined the degree of consistency as *internal consistency*—a measure based on the correlations between different items on the same test. Internal consistency is often measured with *Cronbach's alpha*—a statistic calculated from the pairwise correlations between items—measures the degree of the internal consistency between the several items that propose to measure the same general construct (Cronbach, 1951). The value of Cronbach's alpha (α) may lie between negative infinity and 1, but ranges in value from 0 to 1. Measures that have alpha coefficients that are higher than .70 are considered to possess a high level of internal consistency. Based on Cronbach's definition, calculated alpha values for HIS-EM were considered as a function of the average inter-correlations of HIS-EM dimensions and the number dimensions in the HIS-EM tool while CLASS' Cronbach alpha was considered as a function of the average inter-correlations of CLASS dimensions and the number dimensions in the CLASS tool.

In order to calculate the alpha values and determine the internal consistency of both HIS-EM and CLASS, the remaining 40 videos were scored once independently by either the HIS-EM or the CLASS coders. HIS-EM and CLASS scores for videos that were scored by two coders (*n*=14) were averaged for each dimension to resolve any discrepancies among coders. Averaged HIS-EM and CLASS scores for each video were considered as the final score. All available data was combined to compose the final sample of codes for 54 videos to be analyzed. The results indicated that the Cronbach's alpha correlation coefficient of the HIS-EM was .97 and .86 for CLASS: and revealed that the HIS-EM and CLASS had a high degree of internal consistency.

The Distribution of HIS-EM and CLASS Scores

The final sample of codes for 54 was also analyzed to determine the distribution of the HIS-EM and CLASS scores. Calculating the distribution of the central tendency for HIS-EM and CLASS dimensions and domains (i.e., mean, standard deviation and etc.) will indicate, on average, how teachers scored on HIS-EM and CLASS and how much they varied from one another. The average HIS-EM score (sum of scores for all nine dimensions divided by nine) was 4.46 (on a possible scale of 1 to 7) and ranged from 1.33 to 6.67, with a standard deviation of 1.27. The overall CLASS score (sum of scores for all ten dimensions divided by ten) was 5.69 (on a possible scale of 1 to 7) and ranged from 2.5 to 6, with a standard deviation of .62. Table 25 provides means, standard deviations, minimum, and maximum scores for mean score for HIS-EM and CLASS and each of the HIS-EM and CLASS domain and dimension scored for this study.

Table 25. Descriptive Statistics for HIS-EM and CLASS Domains and Dimensions

Measure	Domain	Dimension	Mean (*SD*)	Range
HIS-EM			4.46 (*1.27*)	1.33-6.67
	What		4.37 (*1.29*)	1.67-7
		Learning Objectives	4.44 (*1.28*)	2-7
		Math Representations	4.63 (*1.33*)	2-7
		Concept Development	4.06 (*1.47*)	1-7
	Who		4.56 (*1.31*)	1.33-7
		Attention to Developmental Trajectories	4.73 (*1.45*)	1-7
		Response to Students' Individual Needs	4.37 (*1.30*)	1-7
		Developmentally Appropriate Learning Formats	4.59 (*1.48*)	2-7
	How		4.49 (*1.33*)	1-6.67
		Planning	4.68 (*1.35*)	1-7
		Student Engagement	4.35 (*1.42*)	1-7
		Establishment of Math Learning Communities	4.45 (*1.38*)	1-7
CLASS			4.83 (*562*)	3.10-6
	Emotional Support		5.75 (*.51*)	4.25-6.50
		Positive Climate	5.73 (*.72*)	4-7
		Negative Climate	6.92 (.27)	6-7
		Teacher Sensitivity	5.50 (*.73*)	3-6
		Regard for Student Perspectives	4.86 (*.93*)	3-6
	Classroom Management		5.41 (*.67*)	3.67-6.67
		Behavior Management	5.50 (*1.00*)	3-7
		Productivity	5.37 (*.68*)	4-6
		Instructional Learning Formats	5.38 (*.84*)	3-7
	Instructional Support		3.04 (*.86*)	1-5.67
		Concept Development	2.81(*.95*)	1-5
		Quality of Feedback	3.34 (*.87*)	1-6
		Language Modeling	2.97 (*1.05*)	1-6

Note: N= 54

Bivariate Correlation Analyses between HIS-EM and CLASS and its Domains

This analysis considers the relationship between two variables rather than analyzing just one independent from the other and allows researchers to investigate the strength of the relationship (range from absolute value 1 to 0) between the variables being analyzed. The stronger the relationship, the closer the value is to 1. Series of bivariate correlation analyses were performed to investigate the strength of the relationship: between HIS-EM and CLASS and between HIS-EM and each CLASS domain (i.e., "Emotional Support," "Classroom Organization," and "Instructional Support"). CLASS domains scores (sum of all dimensions listed under each corresponding domain) and the HIS-EM overall score (sum of all nine dimensions) were moderately correlated (*r*s = .44 to .58), with the strongest relationship occurring between the HIS-EM overall score and the CLASS Instructional Support domain (*r* = .58) (see Table 26).

Table 26. Correlations between HIS-EM and CLASS and CLASS Domains

		HIS-EM (Overall)
CLASS (Overall)		0.54**
	CLASS Emotional Support Domain	0.54**
	CLASS Classroom Organization Domain	0.44**
	CLASS Instructional Support Domain	0.58**

Note: **The correlation is significant at .01 level (two-tailed)

Even though correlation analysis revealed that there is a moderate relationship between HIS-EM and CLASS and its domain scores, it does not provide much information about the direction of the relationship or how much of the variances in CLASS and domain scores can be explained by the changes in HIS-EM score. Therefore, regression analyses were performed to further investigate the relationship between these variables.

Bivariate Linear Regression Analyses between HIS-EM and CLASS and its Domains

Association between HIS-EM and CLASS and its domains scores were investigated by performing series of bivariate regression analyses to determine the direction of the relationship between the HIS-EM and CLASS and its domain scores. First, it was performed between the HIS-EM and CLASS scores, that is, for a particular value of X (i.e., HIS-EM) what value on average do teachers have on Y (i.e., CLASS). Second, it was performed between the HIS-EM and each CLASS domain scores (i.e., “Emotional Support,” “Classroom Organization,” and “Instructional Support”) that is for a particular value of X (i.e., HIS-EM) what value on average do teachers have on Y (i.e., “Emotional Support,” “Classroom Organization,” and/or “Instructional Support”). Results of each of these analyses were performed and reported respectively.

HIS-EM vs CLASS. A bivariate regression analysis was performed utilizing “CLASS” score as the criterion and “HIS-EM” score as the predictor in order to determine if CLASS scores can be predicted as a function of HIS-EM scores. The mathematical equation for this bivariate regression analysis was represented as:

$$Y = \beta0 + \beta1X + \varepsilon$$

Y and X represented the scores for the observed lesson on the CLASS and HIS-EM measures respectively, and the parameters $\beta0$ and $\beta1$ were constants describing the functional relationship in the observed lessons. The value of $\beta1$ identified changes in the

CLASS score expected for every unit changed in the HIS-EM score (representing the slope). The values of $\beta 0$ identified an adjustment constant due to scale differences in measuring HIS-EM and CLASS (the value of Y when X = 0). ε (Epsilon) represented an error component for each observed lesson (the portion of CLASS score that cannot be accounted for by its systematic relationship with values of HIS-EM).

Results indicated that the HIS-EM average score (sum of all HIS-EM dimension scores divided by nine) significantly associated with the CLASS average score (sum of all CLASS dimension scores divided by ten), $R^2 = 0.29$, F (1, 52) = 21.50, $p<0.001$, indicating that HIS-EM score is a good predictor of CLASS score. 29% percent of the variation in CLASS score is associated with variation in HIS-EM scores. More specifically, the regression equation for predicting CLASS score (i.e., Y) from HIS-EM score (i.e., X) was found to be Y = 3.5 + 0.27X , meaning, on average, one unit increase in HIS-EM score is associated with a .27 unit increase in CLASS score (see Figure 9).

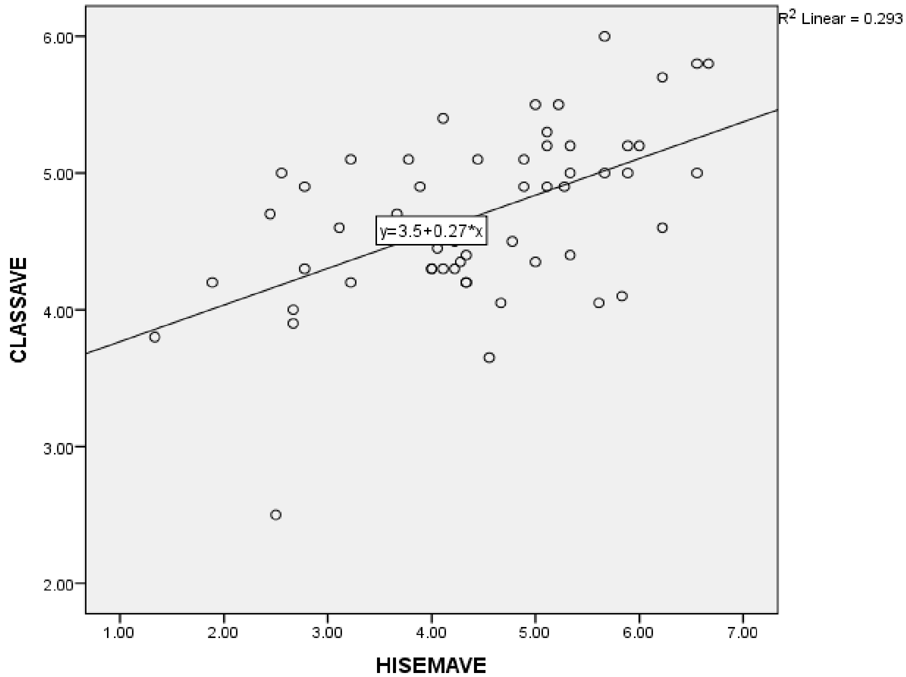

Figure 9. Bivariate Linear Regression Scatter Plot for Average HIS-EM and CLASS Scores

HIS-EM vs CLASS Emotional Support domain. A bivariate regression analysis was performed utilizing "CLASS Emotional Support" domain score as the criterion and "HIS-EM" score as the predictor in order to determine if "Emotional Support" domain scores can be predicted as a function of HIS-EM scores. The mathematical equation for this bivariate regression analysis was represented as:

$$Y = \beta 0 + \beta 1X + \varepsilon$$

Y and X represented the scores for the observed lesson on the CLASS Emotional Support domain and HIS-EM measures respectively, and the parameters β0 and β1 were constants describing the functional relationship in the observed lessons. The value of β1 identified changes in the Emotional Support domain score expected for every unit changed in the HIS-EM score (representing the slope). The values of β0 identified an adjustment constant due to scale differences in measuring HIS-EM and CLASS Emotional Support domain (the value of Y when X = 0). ε (Epsilon) represented an error component for each observed lesson (the portion of Emotional Support domain score that cannot be accounted for by its systematic relationship with values of HIS-EM).

The results indicated that the HIS-EM average score (sum of all HIS-EM dimension scores divided by nine) significantly predicted the CLASS "Emotional Support" domain average score (sum of all dimensions under this domain divided by four), $R^2 = 0.29$, $F(1, 52) = 21.45$, $p < 0.001$, indicating that HIS-EM score is a good predictor of CLASS "Emotional Support" domain score. 29% percent of the variation in CLASS Emotional Support domain score is associated with variation in HIS-EM scores. More specifically, the regression equation for predicting CLASS Emotional support domain score (i.e., Y) from HIS-EM score (i.e., X) score was found to be Y = 4.78 + 0.22X , meaning, on average, one unit increase in HIS-EM score is associated with a .22 increase in CLASS Emotional Support domain score (see Figure 10).

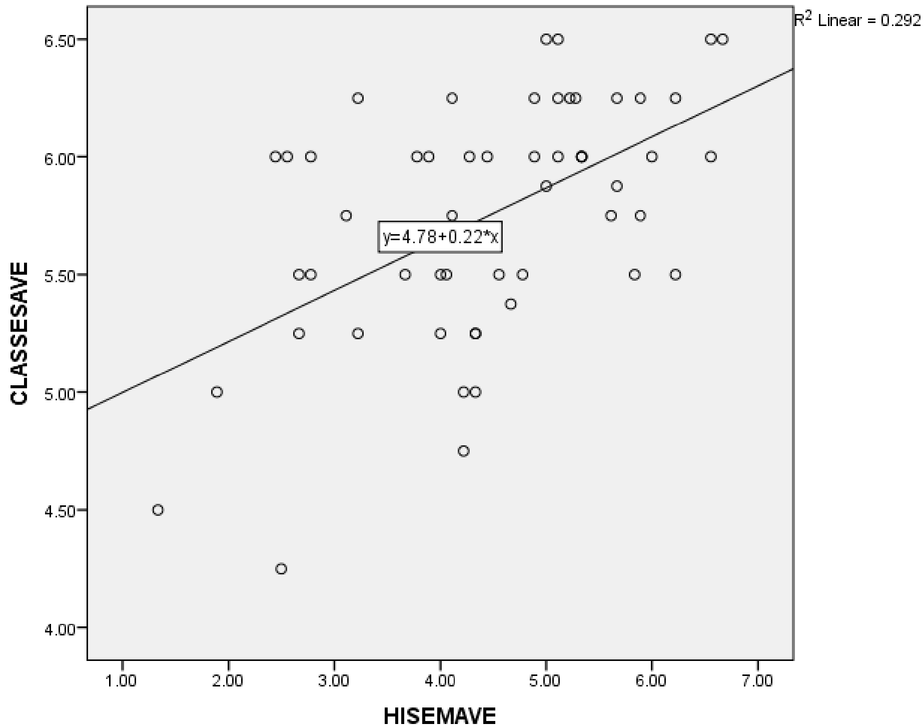

Figure 10. Bivariate Linear Regression Scatter Plot for Average HIS-EM and CLASS Emotional Support Domain Scores

HIS-EM vs CLASS Classroom Organization domain. A bivariate regression analysis was performed utilizing CLASS "Classroom Organization" domain score as the criterion and "HIS-EM" score as the predictor in order to determine if "Classroom Organization" domain scores can be predicted as a function of HIS-EM scores. The mathematical equation for this bivariate regression analysis was represented as:

$$Y = \beta 0 + \beta 1X + \varepsilon$$

Y and X represented the scores for the observed lesson on the CLASS "Classroom Organization" domain and HIS-EM measures respectively, and the parameters β0 and β1 were constants describing the functional relationship in the observed lessons. The value of β1 identified changes in the "Classroom Organization" domain score expected for every unit changed in the HIS-EM score (representing the slope). The values of β0 identified an adjustment constant due to scale differences in measuring HIS-EM and CLASS" Classroom Organization" domain (the value of Y when X = 0). ε (Epsilon) represented an error component for each observed lesson (the portion of Classroom Organization domain score that cannot be accounted for by its systematic relationship with values of HIS-EM).

The results indicated that the HIS-EM average score (sum of all HIS-EM dimension scores divided by nine) significantly predicted the CLASS "Classroom Organization" domain average score (sum of all dimensions under this domain divided by three), $R^2 = 0.19$, $F(1, 52) = 12.47$, $p<0.001$, indicating that HIS-EM score is a good predictor of CLASS "Classroom Organization" domain score. 19% percent of the variation in CLASS Classroom Organization domain score is associated with variation in HIS-EM scores. More specifically, the regression equation for predicting CLASS "Classroom Organization" domain score (i.e., Y) from HIS-EM score (i.e., X) score was found to be Y = 4.37 + 0.24X, meaning, on average, one unit increase in HIS-EM score is associated with a .24 unit increase in CLASS "Classroom Organization" domain score (see Figure 11).

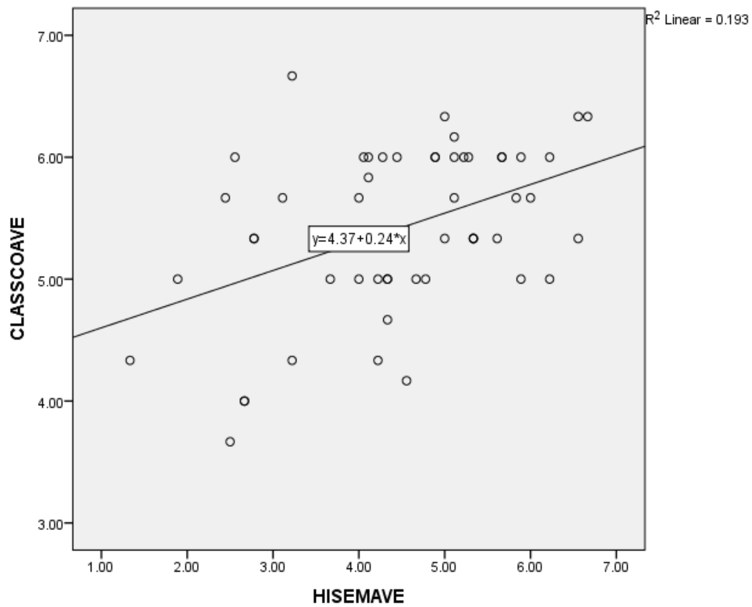

Figure 11. Bivariate Linear Regression Scatter Plot for Average HIS-EM and CLASS Classroom Organization Domain Scores

HIS-EM vs CLASS Instructional Support domain. A bivariate regression analysis was performed utilizing CLASS "Instructional Support" domain score as the criterion and "HIS-EM" score as the predictor in order to determine if CLASS "Instructional Support" domain scores can be predicted as a function of HIS-EM scores. The mathematical equation for this bivariate regression analysis was represented as:

$$Y = \beta 0 + \beta 1X + \varepsilon$$

Y and X represented the scores for the observed lesson on the CLASS "Instructional Support" domain and HIS-EM measures respectively, and the parameters β0 and β1 were constants describing the functional relationship in the observed lessons. The value of β1 identified changes in the Instructional Support domain score expected for every unit changed in the HIS-EM score (representing the slope). The values of β0 identified an adjustment constant due to scale differences in measuring HIS-EM and

CLASS Instructional Support domain (the value of Y when X = 0). ε (Epsilon) represented an error component for each observed lesson (the portion of Instructional Support domain score that cannot be accounted for by its systematic relationship with values of HIS-EM).

The results indicated that the HIS-EM average score (sum of all HIS-EM dimension scores divided by nine) significantly predicted the CLASS "Instructional Support" domain average score (sum of all dimensions under this domain divided by three), $R^2 = 0.34$, $F(1, 52) = 26.30$, $p<0.001$, indicating that HIS-EM score is a good predictor of CLASS "Instructional Support" domain score. 34% percent of the variation in CLASS "Instructional Support" domain score is associated with variation in HIS-EM scores. More specifically, the regression equation for predicting CLASS Instructional Support domain score (i.e., Y) from HIS-EM score (i.e., X) score was found to be $Y = 1.29 + 0.39X$, meaning, on average, one unit increase in HIS-EM score is associated with a .39 unit increase in CLASS "Instructional Support" domain score (see Figure 12).

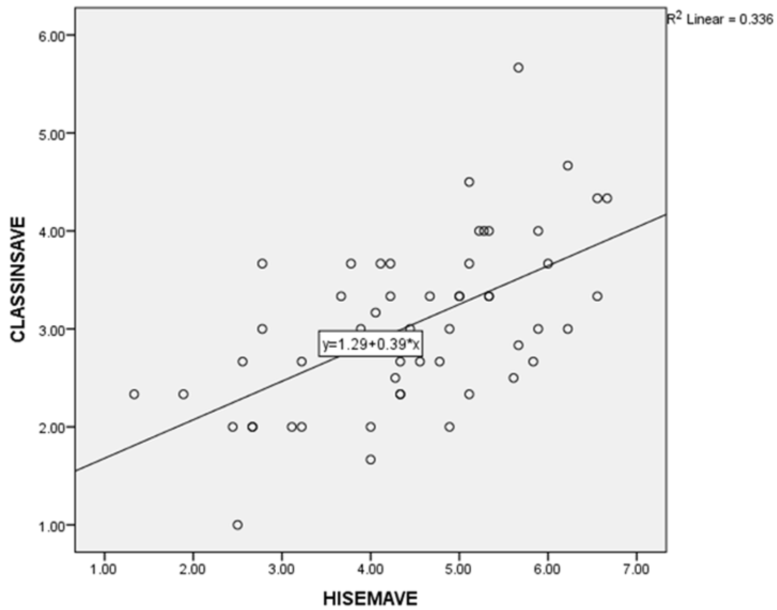

Figure 12. Bivariate Linear Regression Scatter Plot for Average HIS-EM and CLASS Instructional Support Domain Scores

Multiple Regression Analysis between HIS-EM and CLASS Domains

Bivariate correlation and regression analyses revealed that there is a substantial correlation between HIS-EM and CLASS domains. However, these analyses have not revealed when compared to one another, which of the CLASS domain is a better predictor of HIS-EM. A multiple regression is performed as an extension of linear regression to predict the value of HIS-EM based on the value of CLASS domains and to find out which CLASS domain is a better predictor in relation to one another. The mathematical equation for this multiple regression analysis was represented as:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3$$

Y is represented the scores for the observed lesson on the HIS-EM measure. X1, X2, and X3 represented the scores for the lesson on the CLASS domains, “Emotional Support,” “Classroom Organization,” and “Instructional Support” domains respectively.

The parameters β0, β1, β2, and β3 were constants describing the functional relationship in the observed lessons. If X2 and X3 remained the same, the value of β1 identified changes in the HIS-EM score expected for every unit changed in the "Emotional Support" domain score (representing the slope). If X1 and X3 remained the same, the value of β2 identified changes in the HIS-EM score expected for every unit changed in the "Classroom Organization" domain score (representing the slope). If X1 and X2 remained the same, the value of β3 identified changes in the HIS-EM score expected for every unit changed in the "Instructional Support" domain score (representing the slope). The values of β0 identified an adjustment constant due to scale differences in measuring HIS-EM and CLASS domains (the value of Y when X1, X2, and X3 = 0).

The multiple regression model with all three predictors produced $R^2 = 0.41$, F (3, 50) = 11.77, $p<0.001$, indicating that 41% percent of the variation in HIS-EM score is associated with variation in the CLASS domain scores. More specifically, the multiple regression equation for predicting HIS-EM score (i.e., Y) from CLASS domain scores, "Emotional Support," "Classroom Organization," and "Instructional Support" (i.e., X1, X2, and X3 respectively) was found to be:

$$Y = -2.02 + .779X1 + .032X2 + .604X3.$$

Table 27 summarized the descriptive statistics and analysis results. As can be seen, only the CLASS "Instructional Support" domain average score (sum of all dimensions under this domain divided by three) had significant positive regression weights, indicating observed teachers with who have higher scores in this domain were expected to have higher HIS-EM scores, after controlling for the other variables in the model (i.e., "Emotional Support" and "Classroom Organization").

Table 27. Descriptive Statistics for the Multiple Regression Analysis between HIS-EM and CLASS Domain Scores.

Variable	Mean	SD	b	β	*p*
HIS-EM	4.46	1.27			
CLASS Emotional Support Domain	5.75	.51	.779	.313	.120
CLASS Classroom Organization Domain	5.41	.67	.032	.017	.928
CLASS Instructional Support Domain	3.04	.86	.604	.409	.002***

Note: ***The correlation is significant at .001 level (two-tailed)

To confirm these results, a stepwise regression analysis was added to multiple regression analysis to reveal which CLASS domain score is the best predictor of HIS-EM score. Results revealed that "Instructional Support" domain was the single best predictor of HIS-EM (step 1), and "Emotional Support" domain was the next best predictor, only after "Instructional Support" domain was included in the model (step 2). "Classroom Organization" domain did not contribute to the regression model (see Table 28).

Table 28. Descriptive Statistics for the Stepwise Regression Analysis between HIS-EM and CLASS Emotional Support and Instructional Support Domains

Predictors	Mean	SD	b	β	*p*
CLASS Emotional Support Domain	5.75	.51	.813	.327	.000***
CLASS Instructional Support Domain	3.04	.86	.603	.409	.000***

Note: ***The correlation is significant at .001 level (two-tailed)

Conclusion

Overall, results of the bivariate correlation analyses did indicate convergence between the HIS-EM and the CLASS and suggest the existence of the criterion-related validity of HIS-EM with CLASS (meaning both tools are measuring something similar). In particular, convergence between the HIS-EM and the CLASS "Instructional Support" domain was stronger (r = 0.58) than convergence with the CLASS (r=0.54) and its other domains, such as Emotional Support (r=0.54) and Classroom Organization (r= 44). Even though results indicated that the relationship between HIS-EM and "Instructional Support" domain was stronger, r values for others did not drastically differed from one another (ranged from 0.44 to 0.58). Therefore, to better understand how much of the variance in CLASS and its domain scores can be explained by HIS-EM scores and how close and/or far away these scores from each other (indicating the strength of the relationship), series of bivariate regression analyses performed. These analyses revealed that the regression line indicating the closeness of the data points for HIS-EM and CLASS and HIS-EM and CLASS domain (i.e., "Emotional Support," "Classroom Organization," and "Instructional Support") scores to the fitted line actually showed more variations (i.e. .19 to .33). More specifically, the coefficient of determination (i.e., R^2) which in general ranges between 0 (i.e., there is no explanation at all) and 1 (i.e., perfect fit; all variability explained) indicating the strength of the relationship between the two variables was highest for the "Instructional Support" domain (R^2= .34) and lowest for "Classroom Organization" domain (R^2= .19). These results support bivariate correlation analyses and provide a more robust explanation in understanding the relationship between these tools. Multiple regression analysis revealed that "Instructional Support" domain was the single best predictor of HIS-EM and confirmed that HIS-EM scores explained most of the variance in the "Instructional Support" domain and converge more with it, compared to other CLASS domain scores and overall CLASS scores.

Establishing Predictive Validity of HIS-EM

This research question examined the degree to which quality teaching measured by the HIS-EM can be used as an indicator for student learning outcomes in mathematics. The association between quality teaching in mathematics and students' learning gains was explored by analyzing data from participating teachers and students in comparison with group schools within the larger study.

The Distribution of HIS-EM and its Domain Scores

Results suggested that overall teaching quality was medium level (*M*=4.19), ranging from 1.67 to 6.78, with a standard deviation of 1.32 (*N*=73) (see Table 29).

Table 29. Descriptive Statistics of Overall HIS-EM and HIS-EM Domains

		Mean (*SD*)	Minimum	Maximum
HIS-EM (Average)		4.19 (*1.32*)	1.67	6.78
	What (Foundational Knowledge in Mathematics)	4.22(*1.30*)	1.67	7.00
	Who (Knowledge of Young Children)	4.17 (*1.30*)	1.67	6.33
	How (Effective Use of Instructional Support)	4.18(*1.51*)	1.33	7.00

Note: N= 73

Table 30 provides sample sizes, means, standard deviations, minimum and maximum WJ-AP standardized score at each time point. As suggested by the WJ-AP standardized score, assessed students' math performance was lower than the national norm (*M*=100). On average, WJ-AP scores were 95.14 (ranged from 48 to 134) at pre-test and 96.60 (ranged from 49 to 136) at post-test. On average, male students scored higher at both pre-test and post-test compared to female students and Pre-K students scored higher compared to students between kindergarten and 3rd grade (see Table 27).

Table 30. Descriptive Statistics of Students' Mathematical Performance at Pre-test and Post-test by Grade Level and Gender

		WJ III Applied Problems Pre-test (Standardized Score)		WJ III Applied Problems Post-test (Standardized Score)	
Group	N	Mean	SD	Mean	SD
Pre-K	131	99.47	11.40	99.03	11.51
K	160	95.56	11.53	96.58	12.26
1	89	92.88	13.78	93.51	11.45
2	75	93.37	13.73	95.89	15.17
3	91	91.85	14.57	96.74	14.41
Male	259	95.67	13.29	97.56	12.78
Female	287	94.59	12.72	95.73	12.81
Overall	546	95.14	12.99	96.60	12.85

Note: N=546

The Prediction of HIS-EM by Teaching and Professional Development Experiences

Descriptive analyses of the number of years of teaching experience the observed teachers had and the number of math education PD hours they have attended suggested that, on average, observed teachers had 13.7 years of teaching experience, ranging from 1 to 41 years, with a standard deviation of 9.93 (*N*=73). The number of PD hours teachers attended, on average, was 12.1, ranging from 0 to 80 hours, with a standard deviation of 16.63 (see Table 31).

Table 31. Descriptive Statistics of Teaching and Professional Development Experiences

	Mean (*SD*)	Minimum	Maximum
The number of years of teaching experience	13.7 (*9.93*)	1	41
The number of math education PD hours attended	12.1(*16.63*)	0	80

Regression analysis performed to investigate the relationship between teachers' teaching and professional development experiences and their mathematics teaching quality as measure by HIS-EM. The results revealed no statistically significant relationship between commonly used indicators of teacher expertise (i.e., number of years of experience and the number of PD hours teachers attended) and scores on the HIS-EM (observational measure of mathematical teaching quality), R^2= 0.27, F (2, 70) = .973, p= .383.

The Prediction of Students' Math Outcomes by HIS-EM

In order to examine the relationship between quality of mathematics teaching measured by HIS-EM and students' learning gains in mathematics over a school year, three-level hierarchical linear modeling (HLM) analyses (Raudenbush & Byrk, 2002) were conducted by using the HLM program. Hierarchical Linear Modeling (HLM) is a type of regression model often used for analyzing education data sets because they tend to include multiple layers of data that are correlated with one another because they share similar traits (e.g., students from the same classroom and schools are similar in their traits) (Raudenbush & Byrk, 2002). Three-level HLM analysis conducted where students (Level 1) were nested within teachers (Level 2), who were further nested within schools (Level 3), to explore whether students' math outcomes (measured by WJ-AP) was predicted by teachers' quality of teaching in mathematics (measured by HIS-EM), after controlling for school level variances. In this analysis, students' mathematics learning was Level 1 outcome variable and teachers' mathematics teaching quality was Level 2 predictor variable. Model testing is completed in two phases; null model (without predictors) and random intercept and slope model (with predictors at Level 1 and Level 2).

The null model. This model serves as a baseline model with no predictors at any levels (i.e., Level 1, 2, or 3). This model was run first in order to determine the partitioning of variance among the three levels of analysis. The fully unconditional HLM model for WJ-AP test results at post-test used as outcome in 3-level HLM analysis is represented below:

$$\textit{Math Performance at Post-test}_{ijk} = \gamma_{000} + r_{0jk} + u_{00k} + e_{ijk}$$

This model indicates that a student's math performance at post-test is a function of the mean math performance at post-test in the classroom plus some individual variation. The mean math performance at post-in the classroom is a function of the mean math performance across all classrooms at the school level plus some amount of variation between classrooms. The mean math performance at post-test in the school is a function of the math performance at post-test for all schools in the sample plus some amount of variation between schools. Random effects for the intercept at Level 2 (r_{0jk}) and Level 3 (u_{00k}) are the extent to which mean math performance at post-test varied between classrooms and schools, respectively. Analysis of this model revealed $\chi2$ (7) = 11.73, p =.109, and ICC was .01, suggesting that there were not any significant differences in the students' math performance measured by WJ-AP at the school level. Between Level 1 (i.e., student level) and Level 2 (i.e., teacher level), $\chi 2$ (65) = 141.01, $p<.001$, and ICC was .14 suggesting that there were significant differences in students' math performance between classes (within the same school); about 14% of the variance in students' math performance indicated by WJ-AP was between classrooms (i.e., teachers), and about 85% of the variance in students' math performance was between students within a given teacher's classroom.

For this reason, additional predictors to Level 1 and Level 2 were added for further analysis. More specifically, predictors at the teacher level (HIS-EM, Level 2) and student level (pre-test WJ-AP standardized score and students' gender) were added to different models to explore whether, and to what extent, the mathematics performance at pre-test, students' gender, and quality of mathematics teaching measured by HIS-EM explains the differences in math performance at post-test.

The random intercept and slope model. This model predicts the level 1 intercept on the basis of the other grouping or predictor variables. This model was performed after partitioning the variance among the three levels. The WJ-AP pre-test scores (centered around the group) and students' gender (coded dichotomously) were entered to this model as Level 1 predictors of math performance at post-test. The three-level HLM analysis for this model was the following:

$$\textit{Math Performance at Post-test}_{ijk} = \gamma_{000} + \gamma_{100}{*}\textit{GENDER}_{ijk} + \gamma_{200}{*}\textit{WJ-AP-PRE}_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

The addition of gender and math performance at pre-test to this model at Level 1 indicates students' math performance is a function of the mean math performance at post-test in the classroom, plus some effect of gender and math performance at pretest, plus some individual variation. The results indicated that the gender partially significantly predicted the intercept of the level 1 model (*r*= -1.38, p= .07), suggesting that on average boys scored higher than girls in WJ-AP at both pre- and post-test. Also, the pre-test score significantly predicted the slope of the level 1 model (*r*= .66, *p*<.001), suggesting that the higher the pre-test score, the more likely those students performed higher in the post-test as well. In order to further investigate the effects of Level 2 on Level 1 variables, predictors at the Level 2 were added to random intercept and slope model. In this new model was performed in which both WJ-AP pre-test results and students' gender were kept as predictors of math performance at post-test and mathematics teaching quality measured by HIS-EM added as predictor at Level 2. The three-level 3-level HLM analysis was the following:

$$\textit{Math Performance at Post-test}_{ijk} = \gamma_{000} + \gamma_{010}*\mathrm{HISEM}_{jk} + \gamma_{100}*\textit{GENDER}_{ijk} + \gamma_{200}*\textit{WJ-AP-PRE}_{ijk} + \gamma_{210}*\textit{WJ-AP-PRE}_{ijk}*\mathrm{HISEM}_{jk} + r_{0jk} + u_{00k} + e_{ijk}$$

The results indicated that the HIS-EM score did not significantly predict the intercept of the level 1 model (*r*= .559, *p*= .353), suggesting that HIS-EM did not predict students' learning in mathematics after controlling for students' pre-test scores and gender and school level characteristics.

Using one standard deviation above the mean represent high quality mathematics teaching (high scores on HIS-EM), one standard deviation below the mean to represent low quality mathematics teaching (low scores on HIS-EM), and mean score as the average quality of mathematics teaching, observed teachers' HIS-EM scores were categorized as high, low and medium and entered to the model to be analyzed in relation to student outcomes. Even though the overall HIS-EM did not predict students' mathematics learning, the results also suggested that there are varying effects of teachers' math teaching quality on students' mathematics learning. More specifically, teachers who scored high on HIS-EM (one standard deviation higher than the overall mean) more likely to have a positive effect on students mathematics learning at the end of the year (r= .15, p=.027). On the other hand, the effect of students' mathematics performance at the beginning of the school had significantly less effect on their mathematics performance at the end of the school year if they had a teacher who scored average on HIS-EM (r= -.206, p=.001). The negative interaction suggested that while there is a positive relationship between students' pre-test and post-test performance, medium quality of mathematics teaching decreased this relationship. Similar kinds of significant relationships between students' math performance and teachers' math teaching quality were not observed for teachers who score low in HIS-EM (r= .11, p=.16) (see Table 32).

Table 32. Descriptive Statistics for Levels of Teaching Quality in Relation to Student Outcomes

	n	*r*	*SE*	*p*
High HIS-EM	15	.15	.07	.027**
Medium HIS-EM	41	-.20	.06	.001**
Low HIS-EM	17	.11	.07	.160

Note: N= 73
**The correlation is significant at .01 level.

Exploring Teachers' HIS-EM Profile

This research question examined the kinds of teaching profiles existing among early childhood teachers in terms of quality of teaching in early mathematics as measured by HIS-EM. The investigation of the profiles of early childhood teachers' math teaching quality by analyzing; the distribution of HIS-EM scores across teachers, relationship between the HIS-EM domains and how they differed across teachers and whether there clusters of HIS-EM profiles between the teachers.

The Distribution of HIS-EM Scores

To investigate what kind of teaching profiles exist among early childhood teachers in terms of quality of teaching in early mathematics, treatment and comparison group teachers' pre-test HIS-EM scores (from year one of the larger study) were analyzed (*N*=210). The average HIS-EM score (mean across nine dimensions) was 4.06 (on a 1-7 scale; 1 being the lowest and 7 being the highest) and ranged from 1.67 to 6.78, with a standard deviation of 1.24. The medium level results partially supported the hypothesis that average quality of mathematics teaching was mediocre in early childhood settings.

Quality of mathematics teaching was further revealed by three domains of HIS-EM: "what" (foundational knowledge in mathematics), "who" (knowledge of young children), and "how" (effective use of instructional support). The score for each domain was determined by summing up the scores of its dimensions (e.g., "learning objectives," "mathematical representations," and "concept development" for "what" domain) and dividing it by three. As shown in the Table 33, the mean for the level of foundational knowledge in mathematics ("what") 4.04 (*SD*=1.33), knowledge of young children ("who") was 3.96 (*SD*=1.38), and effective use of instructional support ("how") was 4.05 (*SD*=1.52).

Table 33. Descriptive Statistics of HIS-EM Domains

	What (Foundational Knowledge in Mathematics)	Who (Knowledge of Young Children)	How (Effective Use of Instructional Support)
Mean (*SD*)	4.04 (*1.33*)	3.96 (*1.38*)	4.05 (*1.52*)
Minimum	1.33	1	1
Maximum	7	7	7
Percentiles			
25%	3.00	2.67	2.67
50%	4.00	4.00	4.00
75%	5.00	5.00	5.00

Note: N= 210

The Relationship among the Domains of HIS-EM

Correlational analysis was conducted to explore the relationship between the three domains of HIS-EM ("what," "who," and "how"). It was assumed that the 3 dimensions of HIS-EM are moderately correlated. A *correlation coefficient*—the extents to which two variables tend to change together—will be calculated and used to examine the relationships between HIS-EM domains. The coefficient describes both the strength and the direction of the relationship. The most commonly used correlation analyses, Pearson product moment correlation and Spearman rank-order correlation, will be run for the purposes of this study by using correlational analyses. *The Pearson correlation* evaluates the linear relationship between two continuous variables (a change in one variable is associated with a proportional change in the other variable) while *the Spearman correlation* evaluates the monotonic relationship between two continuous or ordinal variables (the change in variables is not necessarily proportional). Both correlation coefficients can range in value from −1 to +1.

The results suggested that there was a significant and positive relationship between the three domains of HIS-EM. The Pearson correlational coefficient (between .907 and .926, $p<.001$) (see Table 34) and Spearman correlations (between .906 and .926, $p<.001$) revealed similar strong correlations suggesting that foundational knowledge in mathematics, knowledge of students, and providing effective instructional strategies are intertwined with each other.

Table 34. Pearson and Spearman Correlations among HIS-EM Domains

Correlations	What (Foundational Knowledge in Mathematics)		Who (Knowledge of Young Children)		How (Effective Use of Instructional Support)	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
What	1	1	.907**	.906**	.914**	.908**
Who	907**	.906**	1	1	.931**	.926**
How	.914**	.908**	.931**	.926**	1	1

Note: N= 210
**The correlation is significant at .01 level (two-tailed).

The Comparisons among the Domains of HIS-EM

Using the sample from Fall 2011 (N=210), a confirmatory factor analyses (CFA) was conducted with the HIS-EM data to test the three-factor model that assumes the 9 dimension reflect three, correlated underlying factors involving the HIS-EM domains of "What," "Who," and "How." The results revealed that the correlations among the What, Who, and How domains indicate that these three domains of HIS-EM are highly related (rs = 0.98 – 0.99) or largely overlapping in terms of what they measure. What and Who share 95.8% of their variance in common; What and How share 95.5% variance in common; and How and Who share 97.7% variance in common.

Clusters of HIS-EM Profiles: Profiles of Early Childhood Teachers' Teaching Quality in Early Mathematics

In order to investigate what kind of HIS-EM profile existed among the observed teachers, two-step cluster analysis[13] was used to identify groups of teachers whose teaching showed similar levels or patterns. This analysis was run on the total sample (*N*=210) on three domains of HIS-EM ("what," "who," and "how"). Overall HIS-EM score was used as the evaluation factor. The results suggested that there were four uniquely profiled groups of teachers whose membership was distributed in a reasonable manner with 15.7%, in cluster 1, 21% in cluster 2, 35.7% in cluster 3, and 27.7% in cluster 4. The summary of the cluster model, including a *silhouette measure of cluster cohesion and separation*, revealed a strong evidence of cluster structure (the silhouette measure average > .5) (Kaufman & Rousseeuw, 1990) regarding interpretation of cluster structures. Based on the profiles to be discussed, the clusters were named as follows: Cluster 1 as" *low,"* Cluster 2 as "*mid-low*," Cluster 3 as "*medium*," and Cluster 4 as "*high."* Domain mean scores and overall HIS-EM scores by cluster were reported in Table 35.

---

[13] Cluster analysis tries to identify homogenous groups of case (i.e., observations).

Table 35. Teachers' Grouping Results from the Two-Step Cluster Analysis

Cluster	Number	%	HIS-EM Mean (SD)	What (Foundational Knowledge in Mathematics)	Who (Knowledge of Young Children)	How (Effective Use of Instructional Support)
Low	58	27.6	2.27 (.40)	2.37 (.49)	2.25 (.49)	2.19 (.51)
Mid-Low	75	35.7	3.86 (.40)	3.90 (.47)	3.84 (.58)	3.85 (.57)
Medium	44	21.0	5.00 (.30)	5.01 (.41)	4.89 (.50)	5.09 (.49)
High	33	15.7	6.12 (.33)	5.98 (.42)	6.01 (.34)	6.36 (.42)

Note: N= 210

As shown in Table 35, 58 teachers out of 210 were in Cluster 1, featured by "*low*" level of quality of mathematics teaching; 75 teachers in Cluster 2, characterized by a "*mid-low*" level of foundational knowledge in mathematics; 44 teachers in Cluster 3, grouped by a "*medium*" level of knowledge of young children; and the rest of the teachers were in Cluster 4 with "*high*" levels of effective use of instructional support.

Overall, the two-step cluster analysis revealed that teachers did have different HIS-EM profiles in regards to the quality of early mathematics teaching they provided. There was only a small portion of the teachers whose HIS-EM scores were higher than that of most of the teachers in the sample (approximately 15%) (see Figure 13).

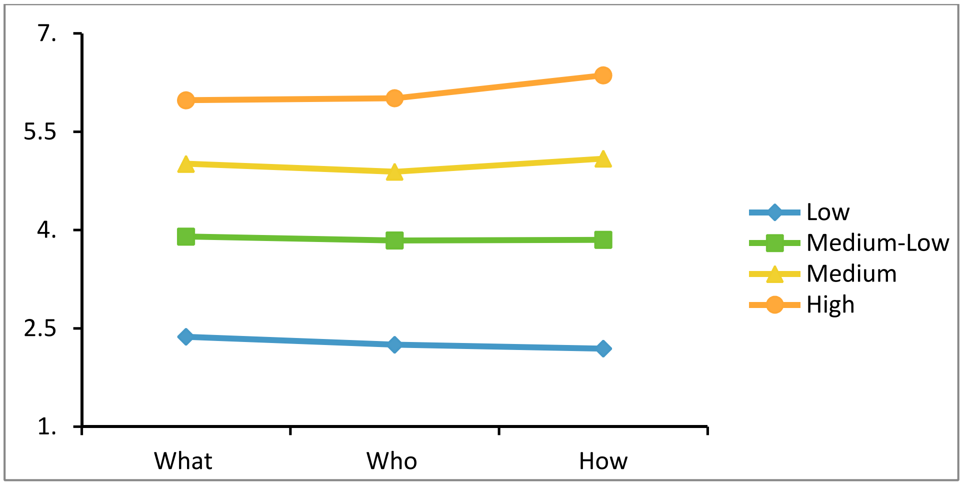

Figure 13. The Profile of HIS-EM by Teachers' Grouping

CHAPTER 5

DISCUSSION

This study investigated the criterion-related validity and the predictive validity of HIS-EM as well as types of early mathematics teaching profiles based on the HIS-EM dimensions among a sample of pre-kindergarten to 3rd grade teachers. By analyzing the observable components of teaching captured by HIS-EM, this study provided new insights for understanding the quality of mathematics teaching. The results suggested that the HIS-EM has the potential to provide a reliable and valid assessment of quality of mathematics teaching in relation to student outcomes, and therefore holds promise as a useful tool in mathematics education research.

In this chapter, the results will be discussed based on the three major questions for the study, and they are:

1. Criterion validity investigation: To what extent will constructs measured by HIS-EM and CLASS converge with or discriminate from one another?

2. Predictive validity investigation: What is the relationship between the quality of mathematics teaching measured by HIS-EM and students' learning outcomes in mathematics?

3. Teaching profile description and analysis: What is the profile of early childhood teachers' mathematics teaching quality as measured by HIS-EM?

Setting the Stage for Further Investigation: Inter-rater Reliability of HIS-EM and CLASS

Before directly addressing the first research question, this section will first explain the need to establish the inter-rater reliability of the observers for the two tools used for the study, namely HIS-EM and CLASS, setting the stage for further investigation. The discussion of the results regarding the reliability establishment for HIS-EM will follow the initial explanation.

As discussed in Chapter 3, reporting inter-observer reliability estimates is especially critical and vital given its potential impact on reliability and validity of the data being analyzed. Specifically, if two raters cannot be shown to reliably rate observed phenomena (rating in similar ways based on the observation protocol), then any subsequent analyses of the ratings given by those raters will yield unreliable and potentially invalid results.

Even though establishing these estimates can be one of the most challenging and more obscure aspects of validating classroom observation protocols due to various and inconsistent methods used by researchers to establish these estimates (Volpe, DiPerna, Hintze, & Shapiro, 2005; Olswang, Svensson, Coggins, Beilinson, & Donaldson, 2006), it is certainly not impossible. There are two important components to note when evaluating the results of an inter-rater reliability analyses. The first is the necessity of establishing the inter-rater reliability estimates based on the methods recommended by the tool developers. Most commonly suggested analyses include percent-agreement, Cohen's kappa, Pearson's r, or Spearman's rho (La Pora, Pianta, & Stuhlman, 2004). Each of these statistics provides a statistical estimate of the extent to which two or more raters rate similarly. For this particular study, percent adjacent agreement (meaning the raters agree within one point of agreement) was calculated to provide information regarding the precise agreement between raters as recommended by the HIS-EM and CLASS manuals and training protocols (above 80% agreement between the coders).

The second important thing to note is that an inter-rater reliability index must be employed that works with the type of rating scale used in the tool. For example, the intra-class correlation (ICC) is a measure of inter-rater agreement that needs to be used when there are 5 or more rating categories or when ratings are made along a continuous scale (e.g., 1 being the lowest and 10 being the highest). Both HIS-EM and CLASS have a rating scale of 1 to 7 (with 1 indicating low and 7 indicating high) which makes it possible to use intra-class correlation (ICC) to provide an estimate for inter-rater reliability. To sum up, for this particular study, inter-rater reliability estimates were established in two ways: percent adjacent agreement and intra-class correlations.

Analysis of the inter-rater agreement estimates using percent adjacent agreement indicated that for HIS-EM, the percent adjacent agreement (equivalent or within one point) across all HIS-EM dimensions was 80.1% and for CLASS, it was 85%. These levels were comparable to the inter-rater agreement reported in the technical manual of

HIS-EM and CLASS (above 80%) and previous CLASS related research (Hamre et al., 2008; Pianta et al., 2008). When examined through intra-class correlations (ICC), based on commonly cited cutoff points by Cicchetti (1994)[14], results suggest that both HIS-EM and CLASS had acceptable degrees of agreement between their coders (generally ranging from "fair" to "excellent"), indicating that HIS-EM and CLASS dimensions rated similarly across the coders. However, there were also cases where the CLASS coders tended to disagree with one another drastically as indicated by lower ICC values for certain dimensions (i.e., "Behavior Management"). These results also mirrored a study in which the researchers were not able to find high level ICCs between CLASS coders (Pianta & Sandilos, 2011). More specifically, researchers rated ICCs as ranging from "poor" to "moderate."

Lower ICCs in this study could be a reflection of the nature of the data coding process. Both HIS-EM and CLASS are developed for use in classroom settings during actual classroom interactions and teaching. Introducing the "videotaped" observation condition rather than completing real-time observations in the preschool and kindergarten classrooms might be a likely reason for the discrepancies observed between the coders. Live observations allow coders to hear conversations and record interactions that would otherwise be inaudible and/or hidden from view in a video recording of the lesson, a circumstance that may compromise reliability. In particular, what the videographer captured or focused on during the course of videotaping might create

---

[14]Cicchetti (1994) suggests commonly-cited cutoffs for qualitative ratings of agreement based on ICC values, with "poor" for values less than .40, "fair" for values between .40 and .59, "good" for values between .60 and .74, and "excellent" for values between .75 and 1.0.

either more limited or richer opportunities of observing evidence for certain domains of CLASS and/or HIS-EM when compared to others depending on what the coder is focusing on recording and coding.

In addition, the format of the math lesson might result in either more limited, or richer opportunities of observing evidence for certain domains of CLASS and/or HIS-EM when compared to others. For example, formal small group instructional time was fairly frequent in observed lessons, therefore limiting opportunities to observe rich examples of certain CLASS dimensions within the "Classroom Organization" domain (i.e., "Behavior Management" and "Productivity"), but not for other dimensions, such as "Instructional Learning Formats." This dimension asks the coders to rate the teacher-child interactions based on the instructional modalities and materials the teacher uses during the course of teaching. Teachers tend to score high in this domain if they provide materials that are of interest to the children. Small group activities by nature tend to focus on specific activities and require the use of materials that will help children engage. Therefore using small group instruction has the potential to boost teachers' scores on this dimension, but not necessarily in others. Furthermore, although the coders were instructed to weight behaviors only across participating students and the teacher, this instruction relies on the subjective judgment of the coder, which is potentially problematic for inter-rater reliability. In particular, some coders might focus on what is happening in the classroom as a whole and code certain dimensions, such as "Productivity," by weighing the behaviors across all children. Other observers might actually only weigh the behaviors across students who are part of the small group activity, and coded accordingly.

Despite the fact that there is room for improvement, both inter-rater reliability analyses revealed that the scores obtained with HIS-EM and CLASS were sufficiently reliable (above 80% percent adjacent agreement between the coders; and ICCs generally ranging from “fair” to “excellent”). Furthermore, the scores obtained for CLASS were close to estimates obtained in other studies (Hamre et al., 2008; Pianta et al., 2008). Having met the psychometric standards for inter-observer reliability between the coders, which is one of the biggest challenges in developing and utilizing classroom observation instruments, HIS-EM appears to be a reliable instrument. Acceptable inter-rater reliability estimates permitted further examination of the HIS-EM’s validity as a next step.

The Criterion-related Validity of HIS-EM with CLASS

Using the CLASS as a criterion measure, the primary goal of this study was to reveal convergent and discriminant patterns between HIS-EM and CLASS and its domain scores. The findings suggest that the HIS-EM and CLASS and its domain scores moderately converged with each other. Four important findings are highlighted below.

First, the results of the study indicate that the HIS-EM produces reliable scores that are correlated in meaningful and predictable ways with CLASS. Specifically, teachers who exhibited higher levels of knowledge of foundational mathematics concepts, understanding of their students’ development and mathematical abilities, and use of effective instructional strategies in mathematics (as measured by HIS-EM) were more likely to provide higher levels of cognitive, behavioral and emotional support to their students (as measured by CLASS) and vice versa.

Second, even though overall CLASS scores were moderately correlated with HIS-EM scores, teachers tend to score differently in CLASS domains. In particular, the scores within the “Emotional Support” (5.75) and “Classroom Organization” (5.41) domains were higher than in the “Instructional Support” domain (3.04). The results were similar to those reported in the literature (La Paro, Pianta, & Stuhlman, 2004; Pianta et. al., 2005; McGuire, Kinzie, Thunder, & Berry, 2016), suggesting that the quality of teacher-child interactions in the present study as reflected by the CLASS can produce comparable results with the quality of teaching observed in other large scale studies where researchers used the tool to measure the general quality of teaching.

Third, the results of the correlational and regression analysis are aligned to highlight the stronger convergence patterns observed between HIS-EM and the “Instructional Support” domain, which is treated as proxy for measuring quality of mathematics teaching practices. Because the HIS-EM captures practices that are specific to mathematics teaching, it has the potential to provide a more fine-grained analysis of instructional quality. For example, “Instructional Support” domain of the CLASS only provides a rating of instructional quality derived from three dimensions (e.g., Concept Development). In contrast, HIS-EM provides a rating of instructional quality derived from nine dimensions (e.g., “Learning Objectives,” ”Planning,” “Student Engagement,” and etc.). In other words, even though the CLASS may be a useful tool in terms of understanding overall instructional quality, the HIS-EM offers a benefit of the multidimensional measure of mathematics teaching quality, which enables the researcher to pinpoint the kinds of strengths that teachers exhibited in math instruction and areas in need of improvement. Additionally, the multidimensional nature of the HIS-EM may help specify the impact of different math instructional practices on child outcomes.

Last but not least, somewhat unexpected in this study were the significant and moderate correlations observable between HIS-EM and the two non-instructional CLASS domains: Emotional Support and Classroom Organization. These two domains were designed to measure general levels of teacher-child interactions. Although at first counterintuitive, the existing literature confirms these findings, noting that fairly close correlations exist between CLASS domains and other content specific instruments. In particular, Walkowiak and her colleagues investigated the correlations between CLASS domains and a measure of standards-based mathematics teaching practices, the Mathematics Scan (M-Scan) (Walkowiak, et. al., 2013). The results of their study revealed that quality of mathematics teaching measured by the M-Scan dimensions were statistically and positively correlated with CLASS "Instructional Support" domain (*r*s ranged from .33 to .48) but also highly correlated with other CLASS domains (e.g., "Emotional Support" and "Classroom Organization") (*r*s ranged between .20 and .42) (Walkowiak, et. al., 2013).

A couple of explanations can be offered to interpret the results of moderate correlations between the HIS-EM results and the two non-instructional domains of CLASS. According to one, the results may indicate that a certain level of emotional and organizational support is necessary in order for teachers to provide quality mathematics instruction as measured by HIS-EM. Early childhood classrooms often are characterized by warmth, nurture, and support for children. These characteristics have long been considered as essential aspects of young children′s education. Therefore, these results may reflect how the instructional supports provided to children are often embedded in or entangled with organization of the classroom and emotional support provided during the course of instruction. Alternatively, it is possible that the association between HIS-EM scores and CLASS's "Emotional Support" and "Classroom Organization" is due to the fact that few of the ICCs that have less than ideal values were grouped under these domains. Low ICCs are indicative of low inter-rater agreement for these domains, which means the raters did not rate some of the dimensions under these domains in a similar fashion. This can be problematic because as the discrepancy between the raters increases, data gathered under these domains becomes less trustworthy and potentially misleading.

In conclusion, the present study provides preliminary but encouraging evidence that the HIS-EM has an established criterion-related validity in relation to CLASS. Knowing that CLASS is not designed to measure quality of early mathematic teaching, the results of this study need to be interpreted cautiously because using a mathematics-specific observation tool as a criterion measure could possibly yield different results. Unfortunately, this was not an option due to lack of available observation protocols designed to focus upon elements of mathematics teaching quality, especially in early childhood classrooms. As such, only limited inferences could be made from the HIS-EM scores in relation to CLASS scores about the extent to which high quality instructional practices are present during the course of mathematics teaching in early childhood classrooms.

Predictive Validity of HIS-EM

The second research question of the study was related to the relationship between the quality of early mathematics instruction as reflected by HIS-EM and student mathematics achievement as measured by the WJ-AP subtest. As measured by HIS-EM, the current study did not reveal a significant prediction of students' mathematical learning over a year after controlling for the impact of students' pre-test scores[15] and gender.[16] The findings of the study also revealed that indirect indicators of teaching experience (i.e., years of teaching and the number of PD hours teachers attended) did not demonstrate any significant association with mathematics teaching quality as measured by HIS-EM. However, the results did find mixed effects of teachers' degree of mathematics teaching quality on students' mathematics learning. Specifically, overall mathematics teaching quality in early childhood classrooms as measured by the HIS-EM was linked to positive child outcomes when the quality of mathematics instruction was identified as "high."

Teaching and Professional Development Experiences

The results showed no statistically significant relationship between commonly used indicators of teacher expertise (i.e., number of years of experience and the number of PD hours teachers attended) and scores on the HIS-EM (observational measure of mathematical teaching quality). Existing research has also shown mixed results on this

---

[15] Students' pre-test score was a significant predictor of their post-test test score. In other words, if a student received a high score at pre-test, they were more likely to receive high score at post-test as well, and vice versa.

[16] On average, boys received higher scores on mathematics achievement tests as compared to girls on both the pre-test and post-test.

matter. For instance, Rockoff (2004) found that the teaching experience of teachers matters, but only up to a certain point. It is generally true that less experienced teachers are less likely to provide quality instruction compared to teachers who have ten to fifteen years' experience. This difference begins to disappear after the less experienced teachers taught about four years (Rockoff 2004; Rivkin, Hanushek, and Kain 2005; Kane, Rockoff, and Staiger 2006). In terms of the relationship between the number of hours teachers participate in professional development in mathematics and higher quality teaching, some found the positive correlations (King & Newmann, 2000) while others reported mixed results (Goldhaber & Brewer, 2000).

Despite the inconclusive results, the current finding is noteworthy because it indicates that the effects of experience, whether measured in years of teaching or hours of professional development, are complex and their association with quality of early mathematics teaching is not linear, at least for this group of teachers. Even though no one would claim that years of teaching experience or professional development services do not contribute to teachers' capacity to provide quality of mathematics teaching, lack of associations might imply that teacher education and professional development programs in early mathematics are not well developed to support teachers. Perhaps the content of these programs and services is not staying up on the latest curricular and pedagogical advances in early mathematics teaching, therefore making it less likely for teachers to deliver quality mathematics instruction regardless of their years of teaching. While what teachers know is the single most important determinant of what students know (Darling-Hammond & Bransford, 2005), it is important for the field to redesign the content of teacher education programs and in-service professional development to ensure the continuity of quality mathematics teaching experiences for all students.

Varying Teaching, Varying Outcomes

The present findings indicate that quality mathematics instruction in early childhood classrooms as measured by HIS-EM does not predict students' learning in mathematics. However, when observed teachers' HIS-EM scores were categorized as high, low and medium (by using minus and plus one standards deviations to represent the varying degrees of mathematics teaching quality) and examined in relation to students' learning gains in mathematics, the results of this study revealed three interesting findings.

First, there was a positive significant interaction between quality of mathematics teaching and students' mathematics achievement at the end of the school year in classrooms where ratings of the instructional quality in mathematics was identified as "high," after controlling for students' pre-test scores and gender. These findings exemplified the significance of higher quality mathematics instruction in facilitating students' mathematics learning. Specifically, teachers who simultaneously exhibited highly sophisticated and developed: (1) understanding of mathematics content, (2) ability to discern the math content based on students' development and learning, and (3) skills in employing a range of strategies to move students along, were able to facilitate their students' mathematics learning. Even though the impact of high quality mathematics teaching on students' learning was rather small and only concerned a subgroup of students, these findings are consistent with other studies indicating the positive effects of high-quality mathematics teaching on student outcomes in mathematics. That is, students of teachers who provide high quality mathematics teaching make more gains in mathematics than their peers in classrooms with lower quality mathematics teaching (Nye, Konstantopoulos, & Hedges, 2004; Kyriakides & Creemers, 2008; Rockoff, 2004).

Second, in classrooms where teachers provided average levels of quality mathematics instruction, there was a negative interaction between quality of mathematics teaching as measured by HIS-EM and students' pre-test and post-test performance. While students' pre-test performance was predictive of their post-test performance, the strength of this relationship decreased when teachers provided mediocre levels of mathematics teaching. This result, however, should not imply that all mediocre quality mathematics instruction is deleterious for students' mathematics learning. Rather, it raises an interesting point which suggests that teachers with average HIS-EM scores may fail to provide consistent level of mathematics teaching and evenly support their students with varying degrees of mathematical abilities. For advanced students, their instructions may not be challenging enough. For students who are behind, adequate support may not be provided.

Third, neither positive nor negative interaction was detected between teachers who provided low quality mathematics instruction and their students' mathematics performance. This finding implies that when teachers failed to; (1) provide students with meaningful mathematics content, (2) provide opportunities for students to engage with and make sense of the mathematics content that is developmentally appropriate, and (3) creating a learning environment conducive to learning mathematics by using effective instructional support in mathematics, no significant relations can be detected between mathematics teaching and learning gains in mathematics. It is not clear why there is no link between the lower quality of mathematics instruction and students' mathematical learning gains. Much remains to be learned about the lower level of quality mathematics teaching and how it affects students' learning in mathematics.

Despite these interesting findings, one lingering question remains unanswered: Why did teachers' mathematics teaching quality envisioned in the HIS-EM not correspond to student achievement gains? One reason for the lack of the association could be the need for more data about students. There are multiple factors (e.g., parents, tutors, and the availability learning materials, classroom size) affecting students' learning outcomes besides the quality of instruction (Koretz & Hamilton, 2005), and the existence of these influences on students' learning make it more difficult to test the relationship between the ratings of quality teaching and student outcomes (Sass, 2008). Such information about students, which can have a potential effect on their mathematics learning, was not collected in this study. Thus, further multifaceted data about students is needed in order to determine how mathematics teaching quality influences student outcomes.

Another reason for the lack of relationship between teachers' instructional practices and student outcomes could involve the state of early mathematics teaching in early childhood classrooms. As discussed in Chapter 2, research reveals that young children come to school with a wealth of mathematics knowledge and that, regardless of their background, all students would benefit from a challenging mathematics education (Claessens, Engel, & Curran, 2014). For example, a study of kindergarten classrooms found that a disparity exists between mathematics teaching and students' abilities: often, teachers spent significant time on mathematics concepts, such as counting and shapes, which most students had already mastered (Engel, 2013). It is a possibility that the majority of the observed teachers' understanding of their students' abilities in mathematics and of what they need to learn might be misaligned with their students' actual abilities and needs. Such misalignment would make it more difficult to test the mathematics teaching quality as measured by HIS-EM in relation to student outcomes, because the tool's framework is based on teachers' understanding of the mathematics content and ability to introduce math concepts that are aligned with their students' development and needs through the use of instructional strategies.

Taken altogether, the results indicated that the interactions between quality of mathematics instruction and the relationship between students' pre-test and post-test math performance were not consistent with regard to the degree of their teaching quality. When there was a statistical impact of teachers' instructional quality on students' learning, the impact was rather small and only concerned a subgroup of students who were taught by high quality teachers. It is also worth noting that no significant relationship was found between low quality mathematics teaching and students' learning gains in mathematics. Preliminary evidence supporting predictive validity of HIS-EM produced mixed results and made it difficult to capture and reveal clear linkage between quality of mathematics teaching and students' outcomes in mathematics across the whole sample of students.

A Profile of HIS-EM among Early Childhood Teachers

As part of an effort to understand the characteristics of early childhood teachers' mathematics teaching quality, a total of 210 pre-kindergarten to 3rd grade teachers were observed with HIS-EM as they taught mathematics lessons. Observed lessons were documented, assessed, and analyzed according to HIS-EM's dimension and domain indicators to investigate what kind of mathematics teaching profiles exist among early childhood teachers.

The descriptive results revealed that the quality of mathematics instruction varies considerably among early childhood teachers as measured by HIS-EM. Some teachers are identified as delivering high quality mathematics instruction because they provide students with opportunities to fully and purposefully engage in deepening their understanding of important mathematics concepts, whereas others are rated as far lower in quality, because their mathematics teaching is very procedural in terms of content emphasized and instructional strategies used. In such situation, learning conceptual mathematics was unlikely, if not impossible. As a whole, observed quality of mathematics teaching was revealed to be mediocre. That is, teachers displayed rather basic knowledge of foundational mathematics concepts (i.e., what), limited understanding of young children's typical learning pathways in mathematics and diverse students' learning needs (i.e., who), and occasional use of instructional support (i.e., how).

The "What" of Quality Mathematics Teaching: Knowledge of Foundational Mathematics Concepts

More than anything else, the literature on mathematics instruction indicates that having a sound understanding of mathematics plays a crucial role in early childhood teachers' ability to communicate mathematics concepts in a meaningful way and to help children make connections and develop their own mathematical ideas (Ball, Lubienski, & Mewborn, 2001; Ma, 1999; Copley, 2004; Ginsburg et al., 2008; Rudd et al., 2008). However how sound mathematics knowledge can be evidenced in mathematics instruction during the course of early mathematics teaching has often been overlooked and is not well studied, especially in early childhood classrooms. The current study investigated the quality of early childhood teachers' knowledge of foundational mathematics concepts and how it surfaces over the course of mathematics teaching as measured through the "what" domain of HIS-EM.

More specifically, the "what" domain examined how mathematics instruction is used to help students comprehend the overarching framework and key ideas of mathematics and construct their own mathematical ideas about the mathematical concepts during the course of mathematics teaching. For young children to grasp foundational mathematics concepts at a high level, mathematics instruction should; (1) encourage them to think critically about the "Big Ideas" of mathematics; (2) provide opportunities for students to demonstrate their math ideas by using mathematical tools and models; and (3) clarify misconceptions and allow students to connect math concepts with their prior knowledge to promote deeper understanding, rather than encouraging students to merely follow math procedures given to them by the teacher, or memorize basic facts or definitions in isolation.

Unfortunately, too few children are exposed to these types of high quality mathematics instruction throughout the early grades. Early childhood teachers in the sample did not demonstrate adequate levels of foundational understanding for high quality mathematics teaching. The mean score for the “what” domain was 4.04 out of 7, and about 85% of the teachers were equal to or below the medium level of understanding, indicating that the majority of the teachers appeared to have basic content knowledge. Although the majority of teachers addressed important mathematics content, more often than not, they failed (1) to incorporate “Big Ideas” of mathematics into their lessons effectively, (2) to articulate connections between mathematical concepts and tools in a way that enabled students to investigate and connect mathematical concepts, and (3) to clarify students’ misconceptions in a way that guided them towards a deeper understanding of the concept under discussion.

Furthermore, the classroom observations showed that some early childhood teachers in the study focused largely on students’ recitative skills in their lessons. They asked students to rote count up to a certain number. Even when they used manipulatives such as unifix cubes to help children stay on track while counting, the teachers failed to help some students use them productively, including correctly using number words in sequence and connecting each number word to one object in order. In other lessons, students spent most of the time playing a mathematics-related game but teachers paid scant attention to the mathematics concepts embedded in the game. Some lessons focused on completing worksheets with no articulation of how the lesson topic was connected to important mathematical concepts. In one case, teacher guided students through the completion of a math worksheet by referring the students to a particular question, telling them to turn to a specific page in their textbook and look for the answer, asking students to read the answer from the book, and then writing the answer on the board. In all these incidences, although teachers set learning goals, these goals were procedural rather than conceptual. The materials provided students with limited opportunities to engage in math learning. Altogether, this level of teaching would lead students toward only partial understanding of mathematical concepts.

It is detrimental for early childhood teachers to have the necessary mathematics knowledge that they need in order to teach the subject effectively (Ball, Lubienski, & Mewborn, 2001; Ma, 1999). Despite its importance however, there is little empirical investigation that has addressed early childhood teachers' understanding of mathematics in relation to their mathematics teaching quality. Even though such studies exist in regards to the upper grades (Ball, Lubenski, & Mewborn, 2001), only limited inferences can be derived from them. By specifically gathering evidence on teachers' understanding of foundational mathematics during the course of mathematics teaching, the current study provided direct empirical evidence suggesting how early childhood teachers' lack of foundational mathematics understanding manifests itself in their teaching and makes it unlikely for them to provide high quality math learning experiences for their students.

The "Who" of Quality Teaching: Knowledge of Young Children

Many early childhood development and education experts emphasize the vital importance of early childhood teachers' familiarity with how young children develop and learn in order to better scaffold the learning process for the child (Vygotsky 1978). Ideally, instruction should correspond with students' development; if the students are actually to learn what is instructed, attention will have to be paid to whether students' instructional experiences are aligned with the trajectory of students' thinking and learning (Clements & Sarama, 2007). Although researchers have described the developmental trajectories and learning progressions for math contents and concepts (Clements & Sarama, 2004; Sarama & Clements, 2009), the degree to which instruction reflects an understanding of young children's development and individual students' learning needs in mathematics is often overlooked and understudied.

The current study examined teachers' understanding of young children ("who" domain of HIS-EM) in terms of their ability to assess what an individual student knows or needs to know about a particular concept and provide scaffolding accordingly, and to use variety of modalities to gain students' interest to further conceptual understanding and learning. For young children to grasp foundational mathematics concepts at a high level, the learning goals and instructional activities of lessons should correspond with the developmental levels of the students and build on students' current level of understanding to move them forward in their math thinking. The use of multiple modalities and learning formats helps build different background knowledge or learning styles of students to engage them in the content learning. It also helps provide opportunities for all students to make conceptual connections about the mathematics concepts to which they are being introduced.

Unfortunately, few children in the present study were exposed to types of early mathematics instruction that correspond with their development and learning. According to the results, early childhood teachers in the study displayed limited levels of understanding regarding how young children approach mathematics and how their mathematics learning can be supported developmentally. Specifically, the mean for “who” domain was 3.96 and about 85% were equal or below the medium level of understanding. Most teachers displayed limited knowledge of the developmental trajectory for the mathematical topic they are teaching and provided some scaffolding that tends to focus on getting the “right” answer and not on building students’ understanding. Although many of the lessons observed were taught in appropriate instructional grouping with an appropriate pacing, they were unfortunately not very productive.

For example, in one lesson the teacher in a 2nd grade classroom used “fill-in-the-blank” questions as a way to “scaffold” students’ learning with an emphasis on getting the right answer. The students were asked to find the right numbers (which numbers to subtract from which number) without connecting these procedures to any meaning. The mathematical content was developmentally appropriate but scaffolding was superficial and the learning format of the lesson emphasized the “completed work” with little concern for evidence or understanding the concept of “taking away.”

According to the "who" domain, teachers also need to understand how individual students learn and how to differentiate their teaching in order to meet the mathematical needs of all students and ensure that no students slip between the cracks. HIS-EM observations revealed that most of the observed teachers displayed knowledge of some of their students' skills and conceptual understandings but not of them. Some of the students were even "left out" of the lesson. For example, in one of the lessons, though the teacher had realized a few students in her small group were not able to recognize numerals higher than 10, no effort was observed during the class to engage them in a way diverging from the techniques used for the rest of the students.

It is vital for teachers to have a deep knowledge base regarding the development of children's mathematical thinking and learning in order to support their students' mathematics learning and understanding. By specifically gathering evidence about teachers' knowledge of young children as presented itself during the course of mathematics teaching, the current study provided direct empirical evidence about early childhood teachers' understanding of young children mathematics development and learning. Specifically, it highlighted early childhood teachers' lack of understanding of their students' development and developmental needs in learning mathematics and how this gap in understanding reflected on their quality of mathematics teaching.

The "How" of Quality Teaching: Effective Use of Instructional Support

On a daily basis, teachers make an abundance of instructional decisions that can either discourage or promote a supportive environment for mathematics learning. In order to effectively develop students' mathematical skills, teachers also need to provide effective instructional support in mathematics. For example, research has demonstrated that student achievement is higher in classes where instructional time is maximized through careful planning (Walberg, 1984). Research also indicates that teachers' questions are crucial in helping students make connections and learn mathematics concepts (Sutton & Krueger, 2002) and promoting a sense of mathematical learning communities (Ball, 1991; Cobb, Yackel, Wood, & Wheatley, 1988) by communicating high expectations for all students and encouraging them to share their ideas and solutions about given problems, are all procedures which are key to quality mathematics teaching.

The current study defined understanding of instructional methods and effective use of instructional support ("how" domain of HIS-EM) in terms of how the teachers interweave the math content and its accompanying pedagogy by planning coherent and conceptual math lessons, engaging children in purposeful mathematical reasoning and inquiry, and fostering a positive disposition towards mathematics during the course of mathematics instruction. For young children to grasp foundational mathematics concepts at a high level, purposeful and thought-provoking math instruction, coupled with opportunities for classroom discussions about students' math-related observations and ideas, must be incorporated into the mathematics lesson.

Unfortunately, only few teachers incorporated the elements of effective use of instructional support into their early mathematics instruction. Early childhood teachers in this study displayed limited use of instructional methods. The mean for "how" domain was 4.05 and about 85% were equal or below the medium level of understanding. Most teachers (1) planned activities that focused on procedures with some connections to underlying mathematical concepts, (2) failed to use questioning effectively to find out what students already know or do not know about a concept addressed to provoke deeper thinking, and (3) occasionally offered encouragement of students' efforts that increases mathematical discussion and risk-taking in sharing ideas.

In general, teachers seemed to struggle to find the balance between establishing a mathematics learning community that encourages students to generate ideas and questions and express their mathematical ideas honestly and openly, and planning pleasant but rigorous math activities. For example, in one of the lessons, the teacher showed genuine enthusiasm for mathematics and had a warm relationship with the students. In terms of mathematical intellectuality however, the mathematical learning community in this classroom was barely existent. Mathematics was presented as combinations of facts and formulas that needed to be memorized by rote. Students rarely received encouragement to share their mathematical ideas with the rest of the students. In some of the other lessons, students were criticized for giving wrong responses. Such a response from the teacher might create a rather hostile learning environment in which it was not acceptable to be wrong while responding to and engaging in the mathematics lesson. Also, the teachers mostly tended to ask closed questions in lessons, tending to evoke only yes/no or "fill- in-the-blank" responses from students. The problem with these closed questions is that it is often difficult to tell if students conceptually understand the content or not and even if they do so, it is still unclear whether others in the class possess a similar level of understanding.

A key facet of effective instructional strategies in mathematics teaching is to help students make sense of mathematics content by connecting the activities of the lessons with effective questioning and ensuring access to opportunities of learning mathematics for all students. By specifically gathering evidence about teachers' use of instructional strategies in providing environments of respect for students' math ideas, questions, and contributions and probing students for elaboration, explanation, justification, or generation of new questions during the course of mathematics teaching, the current study provided direct empirical evidence about early childhood teachers' lack of use of effective instructional support in mathematics teaching.

HIS-EM Domains Collectively Influencing the Quality of Mathematics Teaching

Findings of the present study implied that nine dimensions within the three essential domains of HIS-EM (i.e., what, who and how) are highly related (*r*s = .906 to .931) and largely overlapping in terms of what they measure. Based on Shulman's PCK framework (1986), the conceptual model for HIS-EM claims that for quality mathematics instruction to occur, early childhood teachers need to familiarize themselves with foundational mathematics content (i.e., what), how young children learn in general and specifically in mathematics (i.e., who), and developmentally appropriate teaching strategies to maximize children's mathematics learning and growth (i.e., how). Even though this framework envisions the "What," "Who," and "How" of quality mathematics teaching to be associated with one another, high levels of associations among them was not necessarily expected.

Several reasons might have led to these findings. One possibility concerns the sheer complexity of teaching. HIS-EM acknowledges the complexity of teaching by seeking to illustrate how content knowledge, knowledge of development of young children and their math understanding, and use of appropriate and effective instructional strategies in teaching mathematics required interweaving in practice in order to provide quality mathematics instruction and learning experiences throughout the early years. Rating these indicators of quality teaching simultaneously could very well make HIS-EM domains naturally very difficult to disentangle from one another.

Another reason may be that the definition of HIS-EM domains requires further refinement to ensure that indicators within each domain are clearly defined and distinguishable from each other. It is also equally possible that there are as-yet undiscovered distinct indicators of quality instruction in mathematics that HIS-EM does not necessarily cover. Further refinement of ways to uniquely quantify HIS-EM domains should be explored.

Despite indicating room for improvement, these findings suggest that, rather than emphasizing different components of mathematics teaching over one another, the vision for quality of mathematics instruction should simultaneously emphasize the what, who, and how of mathematics teaching as reflected by HIS-EM. In other words, emphasizing foundational and important mathematics content is critical to quality mathematics teaching. Equally important is the provision of developmentally appropriate mathematics activities that encourage students to participate and engage in problem-solving and provide an environment in which students feel challenged as well as supported.

Variation of Quality in Early Mathematics Instruction

The findings of the study also revealed that most of the observed lessons failed to provide high-quality learning experiences in mathematics to all students. As two-step cluster analysis results suggested only about 15 percent of the Pre-K to 3rd mathematics lessons were classified as high quality, while about 21 percent were medium, 38 percent were medium-low, and 23 percent were low. These results mirrored the results of the Inside Classroom study in which the researchers observed and rated about 360 mathematics lessons in K to 12 and found dramatic variances in the quality of mathematics teaching provided (Weiss, Pasley, Smith, Banilower, & Heck, 2003). More specifically, the researchers suggested that more than half of the observed lesson was considered as low in quality while only 15 percent was high and 27 percent was medium in quality (Weiss, Pasley, Smith, Banilower, & Heck, 2003).

As mentioned previously, three domains of HIS-EM (i.e., what, who, and how) seemed to collectively influence the extent to which the overall quality of mathematics instruction can be enacted during the course of observed mathematics lessons. Consistently, examination of varying degrees of mathematics teaching quality in observed lessons also indicated that how well the teachers showcase the what, who, and how of mathematics teaching as reflected by HIS-EM depends on how well the teacher demonstrated the desirable features of each HIS-EM domain during the course of early mathematics teaching. For example, mathematics instructions of the teachers who were identified as high quality as measured by HIS-EM simultaneously reflected and sheltered the elements of foundational and conceptual mathematics content (Bransford, Brown &

Cocking, 1999; Hamre & Pianta, 2007; Stein, Smith, Henningsen & Silver, 2009), a high level of expertise in understanding how students learn and think about the concepts related to the content that being taught (Emmer & Stough, 2001; Cameron, Connor & Morrison, 2005) and high levels of instructional support in helping students develop an understanding of the mathematical content by providing a challenging yet supportive learning environment (Yackel & Cobb, 1996). On the other hand, low quality mathematics instruction appeared to be lacking intentionality in directing and designing interactions between the content and students. In particular, teachers were not familiar with either the content or its accompanying pedagogy and failed to appropriately challenge, scaffold and extend students' mathematics knowledge and skills in mathematics.

Implications

Although exploratory in nature the findings related to HIS-EM's validation have implications on tool development as well as teacher development. Each is discussed below.

Implications for tool development. Using Pedagogical Content Knowledge (PCK) (Shulman, 1986) as a guiding framework, the HIS-EM constitutes a valuable tool with which to examine the quality of early mathematics instruction. Currently available observation tools for measuring mathematics teaching quality lack a theoretical framework or explicit statement about their theoretical bases. By articulating a conceptual framework and constructing a corresponding instrument to measure the quality of mathematics teaching in early childhood classrooms, HIS-EM delineated components that are essential to quality mathematics teaching. Built on the PCK framework, HIS-EM is designed to offer a language and tool that both promotes and measures the quality of early mathematics teaching.

Findings also imply that the HIS-EM observation rubric provided reliable estimates of various degrees of mathematics teaching quality. A review of existing observation measures in the field of early mathematics education has indicated the need to develop more observation-based instruments to reliably evaluate the quality of early mathematics teaching in classrooms. This study is significant in meeting that need. Specifically, HIS-EM has the potential to be a valuable tool that helps researchers to better understand the range of quality of mathematics teaching existing in early childhood classrooms by detecting various degrees of mathematics teaching quality. In each of the three studies presented earlier, HIS-EM scores were able to detect low, medium, and high quality mathematics teaching observed among early childhood teachers. Because the level is defined in terms of the degree to which a teacher's teaching reflects indicators of quality mathematics instruction, HIS-EM provides specific descriptors of not only where the teacher is at the moment but also how she or he can directionally progress to a higher level of teaching performance.

This study also calls attention to the importance of using content-specific observation measures in examining teaching quality. Nowadays, teachers are held accountable for helping their students meet various content-specific learning standards. While doing so, teachers need to know factors that are influential to different content areas. HIS-EM helps identify such factors specific to early mathematics teaching, focusing on the interactions between teachers' foundational knowledge in mathematics, understanding of young children's learning in mathematics, and effective use of instructional support in mathematics. Compared to a content-general tool such as CLASS, HIS-EM can potentially provide more in-depth information about the quality of mathematics teaching in early childhood classrooms.

Last but not least, the mixed results obtained in this predictive validity study imply that the HIS-EM is still far from perfect. This study relies on the supposition that mathematics instructional quality is related to student learning (Hill et al., 2005; Kersting et al., 2009) and yet does not find a consistent relationship between them. Unfortunately, research has failed to find ways of establishing the predictive validity of several other observation measures by linking student gains with quality of mathematics practice. This study attempted to establish HIS-EM's predictive validity to meet this need, but like the prior research, struggled to find significant results (Sarama & Clements, 2008; Robelen, 2011). While we cannot draw consistent inferences about student learning, findings highlight the importance of providing high quality mathematics teaching in facilitating students' mathematical learning. Finding ways to ensure that high quality mathematics instruction is the norm for all students could be a big first step toward positively impacting students' learning in mathematics.

Implications for teacher development. The study also informs teacher educators and professional development designers about the critical need to improve the preparation and continuing education of early childhood teachers. Specifically, the results portrayed the types of early mathematics teaching profiles existing among early childhood teachers. The commonly observed profile among the early childhood teachers indicates a paucity of high-level mathematics teaching quality and a void of mathematics instruction centered on foundational math concepts that are both developmentally appropriate and lack the use of instructional support in teaching mathematics.

The findings also imply that the vision of high quality mathematics instruction should emphasize the “what,” “who,” and “how” of mathematics teaching simultaneously, rather than advocating one type of pedagogy over another. That is, in order to provide quality mathematics instruction, early childhood teachers need to have sufficient knowledge of the mathematics content they are responsible for teaching. They also need expertise in helping students develop an understanding of that content, including knowing how students typically think about mathematics concepts and, how to determine what his or her students are thinking about those math ideas, and how the available instructional materials can be used to help deepen student understanding (NAEYC-NCTM, 2002). When teachers are able to do this, their instruction is clearer, more focused, and more effective (Kilpatrick et al., 2001). These findings imply that a critical part of the picture may be missing in regards to the current trend in teacher education and professional development which is to focus largely on improving content knowledge in mathematics (Fennema et. al., 1996; Hill & Ball, 2004; U.S. Department of Education, 2010). While supporting knowledge is important, it is equally important to support teachers’ understanding of how young children learn in mathematics and how to provide more effective instructional support in mathematics.

These observed challenges indicate a need for further training of early childhood teachers. This need requires more than simple access to materials, textbooks and instructional materials in early mathematics. In order to shift the current state of early mathematics teaching, steps must be taken to ensure that all early childhood teachers have the necessary pedagogical content knowledge in early mathematics in order to deliver high quality early mathematics instruction that supports students' learning of foundational math concepts. Darling-Hammond summarizes this point well by stating that

> Without knowing deeply how people learn, and how different people learn differently, teachers lack the foundation that can help them figure out what to do when a given technique or text is not effective with all students…this requires incorporating subject matter goals, knowledge of learning, and an appreciation for children's development and needs (Darling-Hammond, 2006, p.4).

Current study suggests that professional development services in early mathematics that blend the what, who and how of early mathematics teaching might hold potential in helping teachers improve their day-to-day mathematics activities and interactions with their students, therefore improving mathematics teaching and learning overall.

Limitations and Future Directions

Clearly, the results presented here are promising, yet preliminary. This study is the first step in the development and validation of the HIS-EM observation measure, a tool that focused on understanding the quality of early mathematics instruction as mathematics teaching occurred. Findings provide initial support for the HIS-EM as a reliable and valid observational assessment of quality mathematics teaching. Given that the HIS-EM is a newly developed tool, it is important to acknowledge several limitations that were not previously addressed.

First, the current study involves a sample of public schools specifically catering to the teachers who are working with students from low-income families. Therefore, it is unclear whether the results can be generalized to different populations. Though developed for use across any early childhood setting, it is impossible to discern the extent to which the HIS-EM may be applicable for settings unrepresented in the current sample, such as private schools and licensed daycare centers. Furthermore, the fact that all teachers in this study are highly educated and certified to teach in early childhood classrooms limits the generalizability of the findings to other less credentialed early childhood teachers. Future research should examine the applicability of the HIS-EM in wider array of early childhood classrooms with greater diversity at both child and teacher level.

Second, the criterion-related validity compared the quality of mathematics instruction with data obtained from two different observation measures (the HIS-EM and CLASS) despite the fact that only the HIS-EM was intended to measure the quality of mathematics instruction. Using a content-specific observation measure in mathematics instead of the CLASS may have potentially yielded different results. Moreover, the inter-rater reliabilities for the global rating scale used in this study, while adequate, were not as strong as might be desired and reflected some inconsistencies between the raters

Third, lack of significant associations obtained in predictive validity of the HIS-EM may be a function of the data collection procedures used and decisions made both at the student and teacher level. This study acknowledges that “standardized achievement tests, in particular, are exceedingly blunt instruments for measuring what students might learn in a given year from a given curriculum" (NRC, 2001, p. 479), and standardized test scores do not always reflect students' actual state of knowledge and abilities (Erlwanger, 1973; Schoenfeld, 1988). It is possible that even though WJ-AP is a standardized and commonly used measure to test students' math achievement, it only provides a snapshot of student achievement at a particular point in time and with limited content coverage (e.g., restricted topics, usually only number). Using outcome tools that measure students' mathematics learning in different mathematics content areas might yield stronger and more consistent results. Furthermore, all HIS-EM data was collected in single-day observations in each teacher's classroom. Unfortunately, single-day observations may not necessarily reflect teacher practice across the entire school year. Synthesis of multiple observation cycles could reveal the true relationship between quality of mathematics and instruction and student achievement that was unable to be detected in this data set.

Fourth, it is also possible that there may be other contributors to students' scores that account for additional variance amongst students' learning gains in mathematics and were not measured either by HIS-EM or the WJ-AP subtest. Future research should also examine how multiple observations within a short timeframe impacts the reliability of the estimates of quality of mathematics instruction in relation to student outcomes in mathematics. For example, most of the curriculums used in the school settings are often organized in units, which tend to change bi-weekly, if not weekly. By observing and documenting multiple mathematics lessons across consecutive days within a single school week or two weeks, researchers can gain a more comprehensive understanding of mathematics instruction and therefore may be able to measure quality mathematics instruction more reliably.

Last but not least, even though this study's findings suggest HIS-EM as a new type of research tool can contribute both to theoretical understanding of quality early mathematics teaching and measurement of quality early mathematics instruction, additional validation of this observational instrument is clearly needed. The conceptual framework for HIS-EM assumes that teacher's understanding of foundational mathematics and students' learning in mathematics and effective use of instructional support in mathematics affect student learning to the extent to which they are reflected on early mathematics instruction. If teacher's understanding of the what, who and how of quality mathematics needs to be translated through instruction to benefit students, the HIS-EM measure, if valid, must predict student outcomes. Additional empirical evidence supporting this theory would help to advance the development of HIS-EM as a measure that can be used to further investigate the complex relationships between quality of early mathematics teaching and student learning.

Conclusion

Many children in the United States lack the opportunity to develop the mathematical proficiency that sets the foundation upon which future learning and success is built. Too many children not only start behind, they continue to lag behind in early mathematics achievement, even after they start school (Clements, 2011). Growing body of evidence indicates higher-quality early mathematics instruction could reverse these trends (Rivkin, Hanushek, & Kain, 2005; Starkey, Klein, & Wakeley, 2004; Weiss &

Pasley, 2004). Unfortunately, most of the early childhood teachers are not equipped to provide children with the kinds of high quality early mathematics instruction and stimulation that they need to learn foundational math concepts. Studies have revealed current issues in the early childhood education field that could withhold teachers form providing high quality mathematics instruction: (1) teacher misconceptions about early math (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006; Varol et al., 2012); and (2) teachers′ lack of knowledge and confidence in teaching mathematics (Ginsburg, Lee & Boyd, 2008; Sarama, DiBiase, Clements & Spitler, 2004).

The typical methods to address these issues in order to improve early mathematics instructional quality have been to develop and publish standards for what students should learn. Common standards and principles that set the stage for early mathematics teaching are without a doubt necessary, but they do not magically translate themselves in to quality teaching and learning experiences. It is the teachers who embrace these standards and provide quality of early mathematics instruction are the ones that make the difference. Therefore, no improvement can be expected to be accomplished without direct attention to the practice of early mathematics teaching and without truly understanding what is happening during the course of mathematics teaching.

Unfortunately, the field of early childhood education and research has also been hampered by the lack of reliable and valid tools for measuring quality of mathematics instruction. Developing the High Impact Strategies in Early Mathematics (HIS-EM) to measure early mathematics teaching quality represents a beginning contribution to this effort. The vision of mathematics teaching that guided this study is based on Pedagogical

Content Knowledge (PCK) framework put forward by Shulman (1986); and claims that for quality mathematics instruction to occur, early childhood teachers need to familiarize themselves with foundational mathematics content (i.e., what) and the ways in which young children learn, specifically in terms of mathematics (i.e., who), and adopt developmentally appropriate teaching strategies to maximize children’s mathematics learning and growth (i.e., how). HIS-EM was designed as an observational measure to document and assess the quality of early mathematics teaching in relation to this vision for mathematics instruction.

The development of any new measure is an iterative process that involves establishing several stringent psychometric properties. Several studies were done to establish various psychometric estimates of the HIS-EM. The compilation of validity evidence and the calculation of reliability coefficients indicate that indicate the HIS-EM shows promise as an observational measure of early mathematics teaching quality. More specially, the study revealed promising evidence to support conceptualization of HIS-EM as a measure of instructional quality. Even though mathematics instructional quality as measured by HIS-EM did not reveal significant prediction of student outcomes, positive significant interaction between high quality mathematics teaching and students’ learning outcomes in mathematics, indicated the vital importance of providing high quality mathematics instruction in order to lead positive learning outcomes in mathematics.

Most notable, the findings of the study also revealed the characteristics various degrees of instructional quality that exist in early childhood classrooms. In particular, mathematics instruction that is judged to be high quality generally shares a number of key elements. Not only they are based on clear and conceptual mathematics learning goals, but they also provide opportunities for students to productively struggle with that content and make sense of the math content by meeting the students where they are at developmentally and providing mathematics learning environments that are simultaneously respectful and challenging of students. In contrast, mathematics instructions judged to be low in quality are characterized by procedural learning goals; learning environments that are lacking in rigor and developmentally appropriate expectations from students; and limited to none existence instructional focus on student understanding and sense-making in mathematics.

When considered in light of the fields' substantial attention to issues of quality of early mathematics teaching and how best to promote students' early mathematics understanding and learning, this study goes a considerable distance in ascertaining which factors indicate the quality of mathematics teaching. While the study will help to contribute to the literature on how to measure early mathematics instruction, more research is needed. If America's low level of mathematics achievement is ever to be interrupted, and if American students are to ever have a chance of succeeding in mathematics, observation measures of early mathematics teaching should continue to seek to understand and identify the characteristics early mathematics instruction that lead to high quality teaching and learning experiences in mathematics. Identifying which types of early mathematics instructions are associated with which developmental outcomes and for whom reflects the sophisticated and nuanced understanding of quality mathematics teaching that is needed to serve the diverse needs of students in America's system of education. Therefore, future work is needed to establish the extent to which the HIS-EM can be: an effective tool for improving teachers' understanding of what, who and how of quality mathematics teaching based on the PCK framework; a source of data to examine relationships among quality of early mathematics instruction and other teacher characteristics; and a predictor of students' learning gains in mathematics.

APPENDIX A

OPERATIONAL DEFINITION OF HIS-EM INDICATORS

Domain	Dimension	Indicator	Operational Definition
What--Knowledge of Foundational Mathematics Concepts	Learning Objectives	Clarity	Learning objectives are clear.
		"Big Ideas"	Learning objectives reflect conceptual understanding and important learning.
		Integrates with prior knowledge	The teacher integrates the lesson with prior knowledge.
		Reorientation statements	The teacher effectively focuses students' attention toward the purpose of the lesson.
	Mathematical Representations	Words and Gestures	Mathematical words and gestures are used frequently and correctly to illustrate concepts.
		Tools	Mathematical tools enable students to investigate concepts and represent their ideas. Connections are made between tools and mathematical concepts.
		Models	Mathematical models are accurate, varied, and help students make connections between concepts.
	Concept Development	Accuracy	The teacher displays deep, connected content knowledge.
		Anticipates common student misconceptions	The teacher anticipates common student misconceptions and successfully clarifies concepts for students.
		Deeper understanding	The lesson leads students to a deeper understanding of the concept.
		Concluding statements	The teacher concludes the lesson by summarizing mathematical concepts and helping students generalize their understanding.
Who--Knowledge of Young Children	Attention to Developmental Trajectories	Typical mathematical development by topic	The teacher displays knowledge of the developmental trajectory for this mathematical topic.
		Scaffolding	The teacher consistently provides scaffolding that builds students' understanding within their mathematical zone of proximal development.
		Using student error	The teacher is consistently responsive to students who make errors and uses "wrong" answers to deepen students' understanding.
	Response to Student's Individual Needs	Differentiation	The teacher displays knowledge of all students' skills and conceptual understanding, including those with special needs. The lesson is differentiated to support all students.
		Monitors student work	The teacher consistently monitors student work and looks for evidence of learning.
			The teacher successfully adjusts the lesson in

		Flexibility	response to students' needs.
		Informal Assessment	Informal assessment is focused on conceptual understanding and process. There is evidence that the teacher has assessment criteria in mind that guides observation and/or documentation.
	Developmentally Appropriate Learning Formats	Grouping	The instructional grouping is appropriate and productive.
		Pace	Pacing of the lesson is appropriate for the students and productive.
		Variety of Modalities	The teacher uses a variety of modalities to effectively interest students and gain their active, hands-on participation.
		Connections	The teacher often helps students connect mathematics to their own experience, to the world around them, and to other disciplines.
How--Knowledge of Instructional Methods	Planning	Activity Selection	The activities of the lesson are focused on exploring mathematical concepts.
		Lesson Design	All components of the lesson are mathematically connected and coherent.
		Preparation	The teacher is fully prepared for the activities.
	Student Engagement	Problem Solving	The teacher provides many opportunities that excite students to participate to engage in problem solving.
		Questions	The teacher frequently asks open-ended questions with more than one possible solution/strategy.
		Explanation and justification	The teacher often asks "what, how, why" questions or otherwise solicits students' explanations/justifications.
	Establishment of a Mathematical Learning Community	Attitude toward mathematics	The teacher show genuine enthusiasm for mathematics.
		Expectations	The teacher communicates high expectations for all students and consistently offers encouragement of students' efforts that increase their persistence.
		Regard for Student Perspectives	The teacher is flexible, incorporating students' ideas when appropriate and allowing choices based on students' interests.
		Mathematical Discussion	Mathematical discussion appears to foster a sense of community in which students feel free to express their mathematical ideas honestly and openly.

APPENDIX B

HIS-EM OBSERVATION SHEET

High-Impact Strategies for Early Mathematics

TEACHING OBSERVATION SHEET

Date: ______________________ School:____

Grade:_____________________ Observer:____

Start Time:_________ End Time:_________ Language of Instruction: __________________

Content Strand (check all that apply; circle major focus) ___ Number and Operations ___ Geometry ___ Algebra ___ Data Analysis ___ Measurement
Instructional Grouping (check all; estimated time spent (%)) ___ Whole Group [# students: ___] (____ %) ___ Partner Work (_____%) ___ Small Group [# students: ___] (____ %) ___ Individual Work (____ %)
Brief Lesson Description:
Math Materials Used:

WHAT	**Learning Objectives**	1	2	3	4	5	6	7
	Notes							
	Mathematical Representations	1	2	3	4	5	6	7
	Notes							
	Concept Development	1	2	3	4	5	6	7
	Notes							
WHO	**Attention to Developmental Trajectories**	1	2	3	4	5	6	7
	Notes							
	Response to Students' Individual Needs	1	2	3	4	5	6	7
	Notes							
	Dev. Appropriate Learning Formats	1	2	3	4	5	6	7
	Notes							
HOW	**Planning**	1	2	3	4	5	6	7
	Notes							
	Student Engagement	1	2	3	4	5	6	7
	Notes							
	Establishment of a Mathematical Learning Community	1	2	3	4	5	6	7
	Notes							

APPENDIX C

ABOUT MY TEACHING:

TEACHER DEMOGRAPHIC INFORMATION SURVEY (FALL 2011)

1. Respondent ID What is the confidential ID number assigned to you?

2. How many year have you been teaching?

3. About how many pre-service math education/methods classes have you taken (excluding all college-level math classes such as calculus and statistics)?

4. About how many hours of in-service math education workshops have you taken in the last two years?

5. Please check all of the teaching certificates you have earned (Check all that apply).
    a. Type 04 early childhood teacher certificate [Birth-Grade 3]
    b. Type 03 elementary education certificate [Grade K-9]
    c. Early childhood special education certificate
    d. Bilingual/ESL endorsement
    e. Special teaching certificate [Grades K-12]
    f. Other (please specify)___________________

6. Do you have a bachelor's degree (BA/BS)? If yes, in subject area (major) did you earn your bachelor's degree?

7. Do you have a Master's degree (i.e., M.S., M.S., M.Ed., etc.)? If yes, in what field or discipline (major) did you earn your Master's degree?

8. Do you speak any language(s) other than English? If so, which language?

    a. Language 1:
    b. Language 2:
    c. Language 3:

9. How would you rate your speaking fluency in each of these languages?

10. When you were school age, was the instructional language at school different from the primary language spoken at home?

11. Have you ever taken any pre-service or professional development course specifically targeted for teaching English Language Learners (ELL) students?

12. How many pre-service and professional development courses have you taken that provide training for teaching English Language Learners (ELL) students?

13. How many years of experience do you have working with English Language Learners (ELL) students in a classroom setting?

14. Does your school have any formal policies about supporting students' home language?

15. Does your school provide bilingual instruction for students?

16. Which of the following bilingual instructional practice, if any, does your school support?

    My school supports some other bilingual instructional practice (What?)

17. How many students are in your class?

18. How many of them speak English as their primary language or only language?

19. How many of them speak English as a second language or are English Language Learners (ELL)?

APPENDIX D

ABOUT MY TEACHING:

TEACHER DEMOGRAPHIC INFORMATION SURVEY (FALL 2013)

The purpose of this questionnaire is to gather information to improve professional development. Please answer all of the questions. We appreciate your time.

1. Respondent ID
   What is your confidential ID number assigned to you?

2. How old are you?
   24 and under
   25-34
   35-44
   45-54
   55-64
   65 and over

3. Are you
   Female Male

4. What is your race or ethnicity?
   African-American or Black
   American Indian or Alaska Native
   Asian
   Caucasian or White
   Hispanic or Latino
   Native Hawaiian or other Pacific Islander
   Other (Please specify)

5. How many of each type of math class did you take in High School (if any)?

	0	1	2	3 or more
Algebra				
Trigonometry				
Geometry				
Calculus				
Statistics				
Other				

6. How many of each type of math class did you take in college and graduate school (if any)?

	0	1	2	3 or more
Math Concepts for Teachers				
Math teaching Methods				
Algebra				
Trigonometry				
Geometry				
Calculus				
Statistics				
Other				

7. How many years have you taught the grade you are teaching now?
   Less than 1 year 1-2 years More than 2 years

REFERENCES

Achieve Inc. (2013). *Closing the expectations gap: 2013 annual report on the alignment of state k–12 policies and practice with the demands of college and careers.* Mountain View, CA: Author.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., and Zhang, J. (2013). *The Condition of Education 2013* (NCES 2013-037). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved April 16, 2015 from http://nces.ed.gov/pubsearch.

Armstrong, A., Ginet, L., & Warisi, S. (2012, January). *Exploring early math teaching and learning in community college ECE programs*. Poster session presented at the annual meeting of the Illinois Education Research Council, Tinley Park, IL.

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20,* 427–435.

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.

Bain, H. P., & Jacobs, R. (1990). *The case for smaller classes and better teachers*. Alexandria, VA: National Association of Elementary School Principals.

Bahr, D. L., & De Garcia, L. A. (2010). *Elementary mathematics is anything but elementary*. Belmont: Wadsworth, Cengage Learning.

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *The Elementary School Journal, 90*(4), 449-466.

Ball, D. L. (1991). What's all this talk about "discourse"? *Arithmetic Teacher, 39*(3), 44-48.

Ball, D. L. (1992). Magical hopes: Manipulatives and the reform of mathematics education. *American Educator, 16*(2), 14-18, 46-47.

Ball, D. L., & Bass, H. (2000). Review of The Number Devil by Hans Magnum Enzensberger. *Notices of the American Mathematical Society, 47*(1), 51-56.

Ball, D. L., & Rowan, B. (2004). Introduction: Measuring instruction. *Elementary School Journal, 105*(1), 3-10.

Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved December 1, 2013, from http://epicpolicy.org/ publication/preschool- education.

Baroody, A.J. (2001). Early number instruction. *Teaching Children Mathematics, 8*(3), 154-158.

Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise.* Mahwah, NJ: Erlbaum.

Battista, M. T. (1999). Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom. *Journal for Research in Mathematics Education, 30*, 417-448.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133-180.

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington, DC: NAEYC.

Bodovski, K & Farkas, G. (2007) Mathematics Growth in Early Elementary School: The Roles of Beginning Knowledge, Student Engagement and Instruction. *The Elementary School Journal, 108*(2), 115-130.

Boston, M. (2012). Assessing Instructional Quality in Mathematics. *The Elementary School Journal, 113*(1), 76-104.

Boston, M., & Wolf, M . K. (2006). *Assessing academic rigor in mathematics instruction: The development of the instructional quality assessment toolkit* (CSE Technical Report 672). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved December 16, 2010, from http://eric.ed.gov/PDFS/ED492868.pdf

Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling-Hammond and J. Bransford (Eds.). *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. (1-39). San Francisco: Jossey-Bass.

Bredekamp, S., & Copple, C. (Eds.) (1997). *Developmentally appropriate practice in early childhood programs* (Rev. ed.). Washington, DC: National Association for the Education of Young Children.

Brenneman, K., Boller, K., Atkins-Burnett, S., Stipek, D., Forry, N., Ertle, B., French, L., Ginsburg, H., Frede, E. & Schultz T. (2011). Measuring the quality of early childhood math and science curricula and teaching. In M Zaslow, I. Martinez-Beck, K. Tout & T. Halle (Eds.), *Quality Measurement in Early Childhood Settings* (pp.77-103). Baltimore Maryland: Paul H. Brookes Publishing Co..

Brophy, J. E. (1991). Conclusion to advances in research on teaching, Vol. II: Teachers' knowledge of subject matter as it relates to teaching practice. In Brophy (Ed.), *Advances in research on teaching: Vol. 2 Teachers' subject matter knowledge and classroom instruction* (pp. 347-362). Greenwich, CT: JIA Press.

Brown, E. T. (2005). The influence of teachers' efficacy and beliefs regarding mathematics instruction in the early childhood classroom. *Journal of Early Childhood Teacher Education, 26*, 239-257.

Bursal, M. & Paznokas, L. (2006). Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science. *School Science and Mathematics, 106*(4), 173-180.

Byrne, C.J. (1983). *Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis and discussion of research strategy*. Paper presented at the meeting of the Northwestern Educational Research Association, Ellenville, NY.

Cady J. & Rearden, K. (2007). Pre-service teachers' beliefs about knowledge, mathematics and science. *School Science and Mathematics, 107*(6), 236-245.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology, 43*, 61–85.

Carpenter, T. P. , Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem-solving in elementary arithmetic. *Journal for Research in Mathematics Education, 19*(5), 385-401.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. A. Romberg (Eds.*), Classrooms that promote mathematical understanding* (pp. 19-32). Mahwah, NJ: Erlbaum.

Carr, M., Peters, S. & Young-Loveridge, J. (1994). Early Childhood Mathematics – A Framework. In J. Neyland (Ed.) *Mathematics Education –A Handbook for Teacher*. Vol 1, pp.262-269. Wellington: Wellington College of Education.

Chen, J-Q., & Cerezci, B. (2014, April). *Aligning the Early Math Professional Development with Teachers' Zone of Proximal Development.* Paper session at the annual meeting of the American Education Research Association, Philadelphia, PA.

Chen, J. Q., McCray, J., Adams, M., & Leow, C. (2013). A survey study of early childhood teachers' beliefs and confidence about teaching early math. *Early Childhood Education Journal, 41*(6), 3-12.

Chicago Department of Family and Support Services. (2014). *Head Start/ Early Head Start annual report 2012-13.* Retrieved December 1, 2014, from http://csdgroup.files.cyscopa.com/Uploads/DFSS%202014%20Annual%20Report%20FINAL%20FINAL.pdf

Chung, K.-E. (1994). *Young children's acquisition of mathematical knowledge and mathematics education in kindergarten*. Unpublished doctoral dissertation, Iowa State University, Iowa.

Cicchetti D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6(4)*, 284–290.

Clarke, B., Doabler, C., Baker, S. K., Fien, H., Smolkowski, K. & Chard, D. (2011, February). Early Learning in Mathematics (ELM): Developing observation instruments to explore mediators of student achievement. *In Paper presented at the 2011 Pacific Coast Research Conferenc*e San Diego, CA.

Clements, D. H. (1999). ''Concrete'' manipulatives, concrete ideas. *Contemporary Issues in Early Childhood, 1*, 45–60.

Clements, D.H. (2001). Mathematics in the preschool. *Teaching Children Mathematics*, *7*, 270-275.

Clements, D. H. (2004). Preface. In D. H. Clements, J. Sarama, & A. -M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. ix-xv). Mahwah: Lawrence Erlbaum.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education, 38*, 136-163.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443-494.

Clements, D. H. & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach.* New York, NY: Routledge.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science, 333*(6045), 968-970.

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

Clements, D. H., Sarama, J., & Gerber, S. (2005, April). *Mathematics Knowledge of Entering Preschoolers*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for early assessment in mathematics.* Columbus, OH: McGraw-Hill Education.

Cobb, P., Yackel, E., Wood, T., & Wheatley, G. (1988). Research into practice: Creating a problem-solving atmosphere. *Arithmetic Teacher, 36(1)*, 46-47.

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (RR-39). Philadelphia: Consortium for Policy Research in Education.

Cohen, D.L., & Hill, H.C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teacher College Record, 102*(2), 294-343.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119-142.

Copley, J. V., (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements, J. Sarama, & A.-M., DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 401-414). Mahwah: Lawrence Erlbaum.

Copple, C. E. (2004). Mathematics curriculum in the early childhood context. In D. H. Clements, J. Sarama, & A.-M., DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah: Lawrence Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika. 16,* 297-334.

Cross, C. T., Woods, T. A. and Schweingruber, H. (eds.) (2009). *Mathematics learning in early childhood: Paths towards excellence and equity*. Washington, DC: National Academy Press.

Darling-Hammond, L. (1999). T*eacher quality and student achievement: A review of state policy evidence*. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammnod, L. (2006). Constucting 21st Century Teacher Education. *Journal of Teacher Education, Vol. 57(10)*, 1-15.

Denton, K., and West, J. (2002). *Children's Reading and Mathematics Achievement in Kindergarten and First Grade* (NCES 2002–125). U.S. Department of Education, NCES. Washington, DC: U.S. Government Printing Office.

Doabler, C. T., Baker, S. K., Kosty, D., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (in press). Examining the association between explicit mathematics instruction and student mathematics achievement. *Elementary School Journal*.

Doabler, C. T., & Fien, H. (2013). Explicit mathematics instruction: What teachers can do for teaching students with mathematics difficulties. *Intervention in School and Clinic, 48*, 276–285.

Dufour–Janvier, B., Bednarz, N., & Belanger, M. (1987). Pedagogical considerations concerning the problem of representation. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics (pp. 109-122)*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.

Duncan, G. & Magnuson, K. (2009). The Nature and Impact of Early Skills, Attention, and Behavior, presented at the Russell Sage Foundation conference on Social Inequality and Educational Outcomes, November 19-20.

Duncan, G. J., & Magnuson, K. A. (2011) The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems in Greg J. Duncan and Richard J. Murnane (eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, New York: Russell Sage, 2011, pp. 47-69.

Early, D. M., Barbarin, O., Bryant, D., Burchinal, M., Chnag, F., Clifford, R., et al. (2005). Pre-Kindergarten in 11 States: NCEDL's Multi-State Study of PreKindergarten and Study of Statewide Early Education Programs (SWEEP): Preliminary Descriptive Report.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36(2),* 102-112.

Engel, M., Claessens, A., & Finch, M. (2011). *Teaching students what they already know? The misalignment between mathematics instructional content and student knowledge in kindergarten.* Paper presented at the American Educational Research Association Annual Meeting, New Orleans, Louisiana.

English, L., & Halford, G. (1995). *Mathematics Education: Models and Processes*. Mahwah (NJ): LEA.

Entwisle, D. R. and Alexander, K. L. (1990), Beginning School Math Competence: Minority and Majority Comparisons. *Child Development, 61,* 454–471.

Epstein, A. S. (2003). How planning and reflection develop young children's thinking skills. *Young Children, 58*(4), 28–36.

Epstein, A. S. (2006). *The intentional teacher: Choosing the best strategies for young children's learning.* Washington, DC: National Association for the Education of Young Children.

Erikson Institute's Early Math Collaborative. (2011). High Impact Strategies in Early Mathematics. Unpublished.

Erikson Institute's Early Math Collaborative. (2013). Big Ideas of Early Mathematics. Pearson Education.

Fennema, E. & Romberg, T. A. (Eds.) (1999). *Classrooms that promote mathematical understanding*. Mahwah, NJ: Erlbaum.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Fuson, K., Carroll, W. M., & Drueck, J. V. (2000) Achievement results for second and third graders using the Standards-based curriculum Everyday Mathematics. *Journal for Research in Mathematics Education 31(3)*: 277-295.

Flanagan, K.D., McPhee, C., & Mulligan, G. (2009). *The Children Born in 2001 at Kindergarten Entry: First Findings From the Kindergarten Data Collections of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).* (NCES 2010-005). Washington, DC: National Center for Education Statistics.

Gay, L. R. (1987). *Educational research: Competencies for analysis and application* (3rd ed.). Columbus, OH: Merrill.

Geary, D.C., Hoard, M.K., & Hamson, C.O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74*, 213–239.

Geist, E. A. (2000). Lessons from the TIMSS videotape study. *Teaching Children Mathematics, 7(3),* 180.

Gelman, R. 2000. The epigenesis of mathematical thinking. *Journal of Applied Developmental, Psychology, 21*(1), 27-37.

Ginsburg, H. (2009). *Early Mathematics Assessment System* (EMAS). NY:Author.

Ginsburg, H. P., Jang, S., Preston, M., Appel, A., & VanEsselstyn, D. (2004). Learning to think about early childhood mathematics education: A course. In C. Greenes & J. Tsankova (Eds.), *Challenging young children mathematically* (pp. 40-56). Boston, MA: Houghton Mifflin.

Ginsburg, H., Lee, J., & Byod, J. (2008). Mathematics education for young children: What it is and how to promote it. *Society for Research in Child Development, Social Policy Report, 22,* 3-22.

Goldhaber, D. D., and Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Education Evaluation and Policy Analysis 22,* 129-45.

Graeber, A., & Tirosh, D. (2008). Pedagogical content knowledge: A useful or an elusive notion? In P. Sullivan (Ed.), *Knowledge and Beliefs in Mathematics Teaching and Teaching Development*. Amsterdam, The Netherlands: Sense.

Graham, T.A., Nash, C., & Paul, K. (1997). Young children's exposure to mathematics: The child care context. *Early Childhood Education Journal, 25*(1), 31-38.

Griffin, S. (2004). Building number sense with Number Worlds: A mathematics program for young children. *Early Childhood Research Quarterly, 19* (1), 173-180.

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.

Grossman, P & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal. 45* (1), 184-205.

Hamilton, L., McCaffrey, D., Stecher, B., Klein, S., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1–29.

Hamre, K. B., Mashburn, A. J., Pianta, R. C., & Locasle-Crouch, J. (2008). *Classroom Assessment Scoring System, Pre-K: Technical appendix*, Baltimore, MD: Paul H. Brookes Publishing Co.

Hamre, B. K., & Maxwell, K. L. (2011*). Best Practices for Conducting Program Observations as Part of Quality Rating and Improvement Systems*, Research-to-Policy, Research-to-Practice Brief OPRE 2011-11b. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hausken, E. G., & Rathbun, A. (2004). *Mathematics instruction in kindergarten: Classroom practices and outcomes.* Paper presented at the American Educational Research Association, San Diego, CA.

Hiebert, J., Morris,A. K., Berk, D., & Jansen,A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education, 58*, 47-61.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *The Elementary School Journal, 101*(1), 3-20.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 372–400.

Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42*(2), 371–406.

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004) Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*, 11-30.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *The Elementary School Journal*, *105*(1), 11–30.

Jalongo, M. R., & Isenberg, J. P. (2000). *Exploring your role: A practitioner's introduction to early childhood education*. Upper Saddle River, NJ: Prentice-Hall.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approache*s. Los Angeles, CA: SAGE publication.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City* (Working Paper 12155). Cambridge, MA: National Bureau of Economic Research.

Kazemi, E., Lampert, M., & Franke, M. (2009). Developing pedagogies in teacher education to support novice teacher's ability to enact ambitious instruction. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 12-30). Palmerston North, NZ: MERGA.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*(1-2), 172–181.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

King, M. B. & Newman, F.M. (2001). Building school capacity through professional development: conceptual and empirical considerations. *The International Journal of Educational Management, 15(2),* 86-94.

Klein, A., & Starkey, P. (2004a). Fostering preschool children's mathematical knowledge: Findings from the Berkeley Math Readiness Project. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics* (pp. 343–360). Mahwah, NJ: Erlbaum.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology, 42*(1), 59-69.

Kilday, C.R., & Kinzie, M.B. (2009). An Analysis of Instruments that Measure the Quality of Mathematics Teaching in Early Childhood. *Early Childhood Education Journal, 36* (4), 1082–3301.

Koretz, D., & Hamilton, L.S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.531-578). Westport, CT: Praeger.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: validation of the COACTIV constructs. *ZDM, 40*(5), 873–892.

Kyriakides, L. & Creemers, B.P.M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19(2),* 183–205.

LaParo, K.M., Pianta, R. C., & Sttuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal, 104,* 409-426.

Larson, M. J., & Whitin, D. J. (2010). Young children using graphs to build mathematical reasoning. *Dimensions of Early Childhood, 38*(3), 15-22.

Layzer, J. L., Goodson, B. D., & Moss, M. (1993). *Final report volume I: Life in preschool.* Cambridge, MA: Abt Associates.

Learning Mathematics for Teaching (2014). A Coding rubric for mathematics quality if instruction (MQI) 4-point version. Unpublished version.

Lee, J. S. (2004). Correlations between kindergarten teachers' attitudes toward mathematics and teaching practice. *Journal of Early Childhood Teacher Education, 25*(2), 173-184.

Lee, E., Brown, M. N., Luft, J. A., & Roehrig, G. H. (2007). Assessing beginning secondary science teachers' PCK: Pilot year results. *School Science and Mathematics, 107*, 52-68.

Lee, J. S., Ginsburg, H. (2009). Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australasian Journal of Early Childhood, 34*(4), 37-45.

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2005). Mathematical Performance Predicts Progress in Reading Comprehension among 7-year Olds. *European Journal of Psychology of Education 20*(2), 121-137.

Lobman, C., Ryan, S., & McLaughlin, J. (2005). Re- constructing teacher education to prepare qualified preschool teachers: Lessons from New Jersey. *Early Childhood Research and Practice*, *7*(2).

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Malara, N., & Zan, R. (2002). The problematic relationship between theory and practice. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 553–580). Mahwah, NJ: Lawrence Erlbaum Associates.

Malmberg, L., Hagger, H., Burn, K., Mutton, T., & Colls, H. (2010). Observed classroom quality during teacher education and two years of professional practice. *Journal of Educational Psychology, 102*(4), 916-932.

Marks, R. (1990). Pedagogical Content Knowledge: From a Mathematical Case to a Modified Conception. *Journal of Teacher Education*, *41*(3), 3–11.

Mayer, R. E. (1999). *The promise of educational psychology: Vol. 1, Learning in the content areas.* Upper Saddle River, NJ: Prentice Hall.

McCaffrey, D. F., Yuan, K., Savitsky, T. D., Lockwood, J. R., & Edelen, M. O. (2014). Uncovering multivariate structure in classroom observations in the presence of rater errors. *Educational Measurement: Issues and Practice.* Advance online publication.

McCray, J. & Chen, J. Q. (2011). Foundational mathematics: A neglected opportunity. In B. Atweh, M. Graven, W. Secada, & P. Valero (Eds.), *Mapping equity and quality in mathematics education*, (pp. 253-268). New York, NY: Springer.

McCray, J. S., & Chen, J. (2012). Pedagogical Content Knowledge for Preschool Mathematics: Construct Validity of a New Teacher Interview. J*ournal of Research in Childhood Education, 26*(3), 291–307.

McGuire, P.R., Kinzie, M., Thunder, K., & Berry, R. (2016). Methods of analysis and overall mathematics teaching quality in at-risk prekindergarten classrooms. *Early Education and Development, 27(1)*, 89-109.

Mix, K. S. (2001). The construction of number concepts. *Cognitive Development*, *17,* 1345-1363.

Miller L.G. & Hays R. D. (2000). Measuring adherence to antiretroviral medications in clinical trials. *HIV/AIDS Clinical Trials*, 1(1), 36–46.

National Assessment of Educational Progress. (NAEP) (2013). *The Nation's Report Card: Mathematics and Reading 2013*. . Retrieved from http://www.nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf.

National Association for the Education of Young Children (NAEYC) (2002/2010). *Early Childhood mathematics: Promoting good beginnings*. Washington, DC: NAEYC.

National Association for the Education of Young Children and National Council of Teachers of Mathematics. (NAEYC & NCTM) (2002). *Early childhood mathematics: Promoting good beginnings.* Retrieved from http://www.naeyc.org/about/positions/psmath.asp

National Commission on Mathematics & Science Teaching for the 21st Century. (NCMST) (2000). Before it's too late: A report to the Nation from the National Commission on Mathematics and Science Teaching for the 21st Century.

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

National Council of Teachers of Mathematics. (NCTM) (2006). *Curriculum Focal Points for PreKindergarten through 8th Grade Mathematics: A quest for coherence*. Reston, VA: NCTM.

National Council of Teachers of Mathematics (NCTM) (2007). *Mathematics Teaching Today.* Reston, VA: NCTM.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237–257.

Olswang, L. B., Svensson, L., Coggins, T. E., Beilinson, J., & Donaldson, A. L. (2006). Reliability issues and solutions for coding social communication performance in classroom settings .*Journal of Speech, Language, and Hearing Research, 49*, 1048–1071.

Organization for Economic Cooperation and Development (OECD) (2012). *Education at a Glance 2012: Highlights*, OECD Publishing.

Ornstein, A. C., & Lasley, T. J. (2000). *Strategies for effective teaching* (3rd ed.). Boston: McGraw-Hill.

Park, S., & Oliver, J. S. (2008). National Board Certification (NBC) as a Catalyst for Teachers' Learning about Teaching: The Effects of the NBC Process on Candidate Teachers' PCK Development. *Journal of Research in Science Teaching, 45*(7), 812–834.

Paterson DL, Potoski B, Capitano B. (2002). Measurement of adherence to antiretroviral medications. *Journal of Acquired Immune Deficiency Syndrome*, *31*(3), 103–106

Perry, B., & Dockett, S. (2002). Early childhood numeracy. *Australian Research in Early Childhood Education, 9*(1), 62–73.

Piaget, J. (1969). *The child's conception of physical causality.* (M. Gabain, Trans.) London: Routledge & Kegan Paul. (Original work published 1927)

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Pre-K Version*. Baltimore, MD: Brookes Publishing.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109–119.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., et al. (2000). *Reformed teaching observation protocol (RTOP): Reference manual* No. ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3-14.

Porter, A. C., Blank, R. K., Smithson, J., & Osthoff, E. (2005, April). *Place-based randomized trials to test the effects on instructional practices of a mathematics/ science professional development program for teachers.* Paper presented at the annual conference of the National Council on Measurement in Education, Montreal, Canada.

Radziszewska, B., & Rogoff, B. (1991). Children's guided participation in planning imaginary errands with skilled adult or peer partners. *Developmental Psychology, 27*(3), 381–389.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition. Newbury Park, CA: Sage*. (Second.). Newbury Park, CA: Sage.

Richardson, K., & Salkeld, L. (1995). Transforming mathematics curriculum. In Sue Bredekamp & Teresa Rosegrant (Eds.), *Reaching potentials: Transforming early childhood curriculum and assessment* (Vol. 2, pp. 23-42). Washington, DC: National Association for the Education of Young Children.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rockoff, Jonah E. 2004. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review Papers and Proceedings, 94(2),* 247-252.

Romberg, T., & Kaput, J. (1999). Mathematics worth teaching, mathematics worth understanding. In E. Fennema, & T. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates.

Rouse, C., Brooks-Gunn, J., & McLanahan, S. (2005). Introduction to school readiness: Closing racial and ethnic gaps. *Future of Children, 15*, 5-13.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What Large-Scale, Survey Research Tells Us about Teacher Effects on Student Achievement: Insights from the “Prospects” Study of Elementary Schools. CPRE Research Report Series.

Rudd, L. C., Lambert, M. C., Satterwhite, M., & Zaier, A. (2008). Mathematical Language in Early Childhood Settings: What Really Counts? *Early Childhood Education Journal, 36,* 75-80.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sarama, J., & Clements, H. D. (2004). Building blocks for early childhood mathematics. *Early Childhood Research Quarterly, 19*, 181-189.

Sarama, J., & Clements, D. H. (2007). Manual for Classroom Observation (COEMET) — Version 3. Unpublished Version.

Sarama, J., & Clements, D. H. (2010). The mathematical lives of young children. In V. Washington & J. D. Andrew (Eds.), *Children of 2020: Creating a better tomorrow* (pp. 81-84). Washington, DC: Council for Professional Recognition.

Sarama, J., Di Biase, A.M., Clements, D. H.,& Spitler, M. E.(2004).The Professional Development Challenge in Preschool Mathematics. In D. H. Clements, J. Sarama, & A.M. DiBiase (Eds.), Engaging Young Children in Mathematics: Standards for Early Childhood Mathematics Education (pp. 415-446). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Sass, T. (2008). *The stability of value-added measures of teacher quality and implications for teacher compensation policy*. Brief 4. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development, 52*(2, Serial No. 216).

Schoenfeld, A. (1988). When good teaching leads to bad results: The disasters of "well taught: mathematics classes*. Educational Psychologist, 23(2),* 145-166.

Sherin, M. G. (2002). A balancing act: Developing a discourse community in a mathematics classroom. *Journal of Mathematics Teacher Education: 5,* 205-233.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Shulman, L., & Grossman, P. (1988). *The intern teacher casebook.* San Francisco, CA: Far West Laboratory for Educational Research and Development.

Shouse, R. (2001). The impact of traditional and reform-style practices on students mathematical achievement. In T. Loveless (Ed.), *The Great Curriculum Debate: How Should We Teach Reading and Math?* Washington, D.C.: Brookings Institution Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-3428.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*(3), 225-273.

Smith, J., Lee, V., & Newmann, F. (2001). *Instruction and achievement in Chicago elementary schools*. Chicago: Consortium on Chicago School Research.

Spillane, J. P. & Zeuli, J. S. (1999). "Reform and Teaching: Exploring Patterns of Practice" in the Context of National and State Mathematics Reforms. E*ducational Evaluation and Policy Analysis*, 21(1).

Starkey, P., Klein, A. and Wakeley, A. (2004) Enhancing Young Children's Mathematical Knowledge through a Pre-Kindergarten Mathematics Intervention. *Early Childhood Research Quarterly, 19*(1), 99-120.

Starkey P, Klein A, DeFlorio L. (2014). Promoting math readiness through a sustainable prekindergarten mathematics intervention. In: Boivin M, Bierman KL, eds. *Promoting school readiness and early learning. Implications of developmental research for practice*. New York, NY: The Guilford Press; 2014:187-210.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development (Second Edition)*. New York, NY: Teachers College Press.

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

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teacher's beliefs and practices related to mathematics instruction. T*eaching and Teacher Education, 17*, 213–226.

Stigler, J.W. & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership, 61*(5), 12-17.

Stuart, E.A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science*, *25*(1), 1-21.

Sutton, J., & Krueger, A. (Eds). (2002). *EDThoughts: What we know about mathematics teaching and learning.* Aurora, CO: Mid-Continent Research for Education and Learning.

Tudge, J., & Doucet, F. (2004). Early mathematical experiences: Observing young Black and White children’s everyday experiences. *Early Childhood Research Quarterly, 19*, 21–39.

Van Den Heuvel-Panhuizen, M. (1990). ‘Realistic arithmetic/mathematic instruction and tests’ in: Gravemeijer, K., Van den Heuvel, M. & L. Streefland, L. (eds) *Contexts, Free Productions, Tests and Geometry in Realistic Mathematics Education,* State University of Utrecht, The Netherlands, OW&OC, pp. 53 - 78.

Varol, F. & Farran, C. (2006). Early mathematical growth: how to support young children’s mathematical development. *Early Childhood Education Journal*, 33(6), 381-387.

Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2006). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*, 454 473.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Michael Cole, Vera John-Steiner, Sylvia Scribner, & Ellen Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.

Walberg, H. J. (1984). Improving the productivity of America’s schools. *Educational Leadership, 41(8),* 19–27.

Walkowiak, T. A., Berry III, R. Q., Meyer, J. P., Rimm-Kaufman, S. E., & Ottmar, E. R. (2013). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics, 85*, 109-128.

Walls, R. T., Nardi, A. H., von Minden, A. M., & Hoffman, N. (2002). The characteristics of effective and ineffective teachers. *Teacher Education Quarterly, 29*(1), 39-48.

Weiss, I. R. & Pasley, J. D. (2004). What is high-quality instruction? *Educational Leadership, 61*(5).

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research, Inc.

Wilkins, J.L.M. (2008). The relationship among elementary teachers’ content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139-164.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358,* 749-750.

Yackel, E. and Cobb, P. (1996) Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 22,* 390-408.

VITA

Bilge Cerezci was raised in Istanbul, Turkey. Before attending Loyola University Chicago, she attended the Erikson Institute, where she earned a Master's degree in Child Development with Infancy Specialization. She also attended Bogazici University, where she received a Bachelors Early Childhood Education. While at Loyola, Cerezci received the Irving Harris Leadership scholarship and worked as a Doctoral Research Fellow for Early Math Collaborative at Erikson Institute. Currently, Cerezci is a Research Associate at Early Math Collaborative and an Adjunct Faculty at Chicago City Colleges. She lives in Chicago, Illinois.