NEW APPROACHES TO INCREASING EQUITY AND ACCESS TO QUALITY MATHEMATICS INSTRUCTION THROUGH EARLY SCREENING: A VALIDATION OF THE NUMBER SENSE BRIEF (NSB) AND THE TEST OF EARLY NUMERACY (TEN)

by

SARA ELIZABETH WOODRUFF

(Under the Direction of Martha Carr)

ABSTRACT

Kindergarten is the first formal schooling experience for many children, presenting a critical opportunity to develop early mathematics skills. Subsequently, the development of number sense and related mathematics skills is of critical importance. Universal screening for mathematics difficulties provides a tool for identifying students that are at-risk for future mathematics difficulties upon kindergarten entry, and monitoring their mathematics progress throughout the school year. Given that mathematics trajectories are often predicted by initial skill level, early prevention and intervention can greatly improve mathematics outcomes. Furthermore, universal screening may improve equity and access to quality mathematics instruction for children across racial and ethnic groups by identifying areas where additional support is needed. However, there is limited research on the predictive validity of these assessments, specifically the functioning of cut scores, across racial and ethnic groups. The current study is a validation of two such screening assessments, the Number Sense Brief (NSB) and the Test of Early Numeracy (TEN).

INDEX WORDS: Kindergarten, Mathematics, Screening, Assessments, Validity

INCREASING EQUITY AND ACCESS TO QUALITY MATHEMATICS INSTRUCTION THROUGH EARLY SCREENING: A VALIDATION OF THE NUMBER SENSE BRIEF (NSB) AND THE TEST OF EARLY NUMERACY (TEN)

by

SARA ELIZABETH WOODRUFF

BA, Duke University, 2004

MA, University of Georgia, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

NEW APPROACHES TO INCREASING EQUITY AND ACCESS TO QUALITY MATHEMATICS INSTRUCTION THROUGH EARLY SCREENING: A VALIDATION OF THE NUMBER SENSE BRIEF (NSB) AND THE TEST OF EARLY NUMERACY (TEN)

by

SARA ELIZABETH WOODRUFF

Major Professor: Martha Carr

Committee: Jerome Morris
Stacey Neuharth-Pritchett
Paula Schwanenflugel

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2013

DEDICATION

I would like to dedicate this effort to all who believe in education's possibility to alter and shape lives, and particularly those who have dedicated their lives to this critical work. My hope is that educational equity becomes a national imperative.

ACKNOWLEDGEMENTS

I could not have completed this journey without family and friends that have been steadfast in their support and encouragement. I am more than I could ever be without you. To my friends: Beryl, Stephanie, Amanda B., Amanda F., Adeoye, Chris, Nichole, Monique, Sylvia, Heather, Ashlee, Ashley, Nicole, Sarah, Ashton, Eliza, and anyone else I have unintentionally omitted, I hold you all close to my heart. Beryl, we will always remember Fridays at Panera. I am especially indebted to my mother, Sylvia Battle, who has supported me in every way during this education journey—knowing not to ask when I would graduate. Your investment in my education since childhood has forever altered the trajectory of my life. I am grateful for my patient major professor, Dr. Marty Carr; she encouraged me to follow the research and career path that I desire. In addition, I am indebted to Dr. Jerome Morris, whose mentorship and community approach to research has immeasurably enriched my doctoral experience. Additionally, my working and personal relationship with Dr. Paula Schwanenflugel has created everlasting memories and invaluable opportunity. I cannot forget my GCA family for their patience, encouragement, and flexibility, which pushed me to end, when I needed it most.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

CHAPTER

1 Introduction 1

2 Mathematics Development 8

Mathematics and School Readiness 9

Transition to Kindergarten 10

Teacher Attitudes 11

Policy and Early Learning 15

Structural and Economic Factors 17

Number Sense 18

3 Cut off Scores, RTI and ROC Analyses 30

Response to Intervention (RTI) 31

4 Current Study 36

5 Methods 39

Participants 39

Measures 40

6 Data Analysis 45

Signal Detection Theory 45

Additional Analyses 46
Statistical Power 46
Cut Scores 48
7 Results 52
8 Discussion 60
Limitations 64
9 Implications and Recommendations 66
10 Conclusion 68
REFERENCES 69
APPENDICES
A Parental Permission Form 90
B Number line/number magnitude 93

LIST OF TABLES

Page

Table 1: Early Mathematics Assessments and Validity Information..42

Table 2: Descriptive Statistics 52

Table 3: Correlations...54

Table 4: Partial Correlations 55

Table 5: Repeated Measures Analysis of Variance (Within Subjects) 57

Table 6: Repeated Measures Analysis of Variance (Between Subjects) 57

Table 7: ROC Validated Against NSB Recommended Cut Scores 1st Grade 58

Table 8: ROC Validated Against NSB Recommended Cut Scores Kindergarten.........................58

CHAPTER 1

Introduction

The state of mathematics education and achievement in the United States has received increased attention and scrutiny on a national level (Witzel & Riccomini, 2007). This attention is highly warranted given that national evaluations like the National Assessment of Educational Progress (NAEP) confirm that students in the United States continue to lack the requisite skills for mathematics success. The National Association of Educational Progress (2009) revealed no improvement for fourth graders in math since 2007. Of grave concern is the fact that 82% of fourth graders performed at or above basic level; 39% of students perform at or above proficient level. The levels as the defined by the NAEP may prove difficult to interpret given their broad descriptions: “basic denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade; proficient represents solid academic performance. Students at this level have demonstrated competency over challenging subject matter. Advanced represents superior performance” (NAEP, 2009, pp. 9). The NAEP outlines the skills that are associated with each level of performance. Equally, if not more troubling is the gap that persists between White students and African American, Latino, and Native American, and Pacific Islanders. Although African Americans and Latino students experienced improvement in mathematics over previous years, this growth was not adequate enough to begin to eliminate outcome disparities between these groups.

The overall disdainful performance of American students in mathematics may be a consequence of Americans' acceptance of poor mathematics skills and mathematics illiteracy as the norm—an attitude that is not true for reading (Ladson-Billings, 1997). While there is shame associated with the inability to read, there is little shame in poor mathematics proficiency (Ladson-Billings, 1997), and indifference (Paulos, 1988; Ball, Lubienski, & Mewborn, 2001). As a result of American students' lack of exposure and success in higher-level mathematics courses, like calculus, there is a paucity of workers to fill critical jobs, which inevitably will be occupied by skilled foreign workers (Augustine, 2005)—jobs that increasingly require higher levels of mathematics competence (Amit & Fried, 2002). Higher-level mathematics skills are also directly connected to STEM careers, which remain in high demand and command higher salaries (Heckert et. al, 2002). Taking mathematics courses such as Algebra II and Trigonometry is particularly critical for increasing the representation of African American and Latinos in the STEM disciplines (AAAS, 2001). African American students are at greater risk for negative schooling outcomes, so improving their access to quality mathematics instruction may improve their life chances (Ladson-Billings, 1997). Mathematics may be considered the new civil rights battleground (Moses, Kamii, Swapp, Howard, 1989; Ladson-Billings, 1997; Cobb & Hodge, 2002) because algebra is a gateway course to the upper-level mathematics, STEM careers, and better life choices. African American students are less likely to either take and less likely to experience success in Algebra. Mathematics knowledge is a conduit to social access and mobility, thereby resulting in economic and social disenfranchisement when this access is denied (Schoenfeld, 2004).

The performance disparities between Whites and African American and Latino students have been discussed extensively in the academic literature (NAEP, 2009); however, disparity in access to quality mathematics has been addressed only peripherally with regard to young children. Most research has focused on high school and middle school students. The lack of research on such disparities in access for young children has marginalized the effect of race, class, and culture on mathematics learning at the earliest levels. Yet, there is evidence that these factors impact mathematics learning for young children. There is a need for greater mathematics access and reform in instruction at the pre-kindergarten and kindergarten—where modified instruction and intervention can prevent many of the pervasive challenges that arise for students in middle and high school.

In response to the need to identify kindergarten students' mathematics skills, myriad assessments have been created to measure students' early mathematics ability. The skills measured in these assessments, as well as reported validity information, are detailed in Table 1. While there is some diversity in the depth, breadth, and length of mathematics assessments for kindergarten students, there are several core skills that are prevalent across assessments. These include: counting, which includes oral counting and filling in a missing number in a string of numbers; quantity discrimination, which requires the identification of the larger or smaller number in a set of numbers; word/story problems that require simple addition and subtraction; basic addition and subtraction problems, including math facts; number identification; sorting and classification tasks.

The current study examines the validity of two kindergarten mathematics assessments intended to detect students' early mathematics skills and their risk for future mathematics difficulties. This study seeks to compare the predictive validity of the two assessments, the Test of Early Numeracy (TEN) and the Number Sense Brief (NSB). In addition, this study interrogates the use of universal cut scores that is pervasive in assessment and whether or not these scores function equally across race and ethnicity to predict students' risk for mathematics difficulty beyond the kindergarten year into first grade. Cut scores are score(s) used to divide a score scale or other data into categories, in which inferences or decisions are made based on these classifications (Dwyer, 1996). Contrary to popular thought, the passing score [or cut score] for an assessment is not absolute, but rather a point of a defined scale, which is associated with a given performance standard—a minimum performance level (Kane, 1994). The goal of a screening measure [used in an education setting] is to correctly classify children, while minimizing erroneous classifications (DeStefano, 2011)—classifications which are made based on cut scores. In the current study, assessing the development of kindergarten mathematics skills, and identifying "risk" for experiencing future mathematics difficulties are the screening objectives.

The literature review outlines the mathematics development expected in young children at and around kindergarten age. There is a temptation to view mathematics readiness strictly in cognitive terms (with respect to number sense and related skills), as exhibited by mastery of a set of math skills in a given setting (Lee, Autry, Fox, & Williams, 2009; Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006). Yet, differences in development are a result of a confluence of factors including children's social-emotional development, student-teacher interactions, and teacher quality and instruction—in addition to number sense. However, early mathematics cognition, including number sense, is the focus of the assessments in the current study. I will also review short-term memory and speed of processing, and strategy use—all of which are measured directly or indirectly in the assessments administered in this study. There is a particular interest in examining some of the sociocultural challenges faced by African American, Latino, and poor children with respect to mathematics development. These factors will not be directly measured outside of collecting relevant demographic information, but it stands to reason that they impact mathematics achievement in kindergarten and beyond.

Given that kindergarten mathematics achievement is highly predictive of first grade achievement (Jordan, Kaplan, Locuniak, & Ramimeni, 2007; Jordan, Kaplan, Ramineni, & Lociniak, 2009), the early identification of the skills and deficits of young children has become a priority (Fuchs, et al., 2005; Jordan, Kapla, Olah, & Locuniak, 2006; Seethaler & Fuchs, 2010), resulting in the development of mathematics assessments for kindergarten students, which screen for early mathematics achievement (Gersten, Jordan, & Flojo, 2005; Jordan, Kalpan, Olah, & Locuniak, 2006; Seethaler & Fuchs, 2010) . The trajectory for students experiencing severe learning difficulties in mathematics is bleak (Clarke et. al, 2011). Data from the *Early Childhood Longitudinal Study-Kindergarten* (*ECLS-K*) reveal that children with the lowest scores at fall and spring of kindergarten have the lowest scores at fifth grade (Morgan, Farkas, & Wu, 2009; Clarke et. al, 2011), supporting the notion that students that have early difficulties that persist without appropriate remediation, are not likely to experience improvement. This research contributes to our understanding of this issue by validating existing assessments to determine whether the Latino and African American students, (from a rural Title-I school) in this sample demonstrate similar performance to the larger population used to provide norms for this assessment. This research may also demonstrate the need to vary the cut points used to determine risk for failure or difficulty in early mathematics based on the demographics of the sample where these scores are applied.

Kindergarten presents a unique opportunity to assess children's mathematics progress as they matriculate through their first year of formal schooling. Such assessment is critical because although kindergartners are young children, diverse skill levels still emerge, although not immediately detected (Chard et. al, 2008). The assessment of children upon kindergarten entry and throughout the school year, ideally, allows teachers the opportunity to tailor instruction to address students' strengths and weaknesses, catch students that are likely to struggle, and face difficulties into first grade and beyond (Jordan, Glutting, Ramineni, & Watkins, 2010; Jordan, Kaplan, Ramineni, & Locuniak, 2009). And early screening facilitates appropriate intervention for students that continue to struggle through the school year (Clarke, Baker, Smolkowski, & Chard, 2008). The current study examines the predictive validity of several screening measures between kindergarten and first grade.

However, it is not sufficient to understand the mathematics achievement of Latino, African American, and poor students through the lens of core competencies alone. Instead, it is necessary to also consider the sociocultural factors that shape mathematics achievement and mathematics learning. Only in the last decade, has mathematics been seen as more than a cognitive process that is connected to both sociocultural and sociohistorical contexts (Moffatt, Anderson, Anderson, & Shapiro, 2009). In addition to competence with mathematics pedagogy and instruction, teacher resistance may prove a significant structural and social obstacle to improving the achievement of African American students (Rodriguez, 2005; Wells & Serna, 1996; Cockrell, Placier, Cockrell, Middleton, 1999). From a pedagogical approach, many teachers may reject mathematics teaching practices that emphasize constructivism and inquiry-based learning, favoring the more traditional practices of lecturing and assigning individual work (Rodriguez, 2005). Additionally, language and reading skills may affect mathematics achievement. Underdeveloped language skills may interfere with problem comprehension and the overall demands of the task, specifically understanding orally presented problems and verbally expressing their own thoughts (Pappas, Ginsburg, & Jiang, 2003). For example, in a study comparing vocabulary growth of African American and White children using the Peabody Picture Vocabulary Test (PPVT), on average, White children had significantly higher vocabulary knowledge than African American children; White children reached the 40-word level at approximately 50 months of age, whereas African American children did not reach this mark until 63 months—almost a year later (Farkas & Beron, 2004). Increasing numbers of diverse students are now entering the educations system, and narrowly-defined assessment tools may be inappropriate (Kagan & Kauerz, 1997) and fail to adequately capture student knowledge. Standardized assessments have been historically designed by and for well-educated White students (Manly, 2004). A related issue is the prevalence of a tester-subject interaction. When Black and Latino students were administered tests by familiar examiners they performed differentially (Fuchs & Fuchs, 1989; Brooks-Gunn, Klebanov, Smith, Duncan & Lee, 2003), while this was not true for White students. There are numerous other sociocultural factors that may impact student mathematics achievement, but these factors may begin to expose the complexity of these issues. This study considers these factors by focusing the study on an overwhelmingly African American and Latino sample that attends a Title I school in order to validate these assessments that are normed on a majority White population.

CHAPTER 2

Mathematics Development

The foundation for success in mathematics is laid in the early years of schooling. The kindergarten year is critical as learning in kindergarten impacts first grade success; children who are more successful in first grade experience greater success in the later grades (Alexander & Entwisle, 1988; Entwisle, Alexander, & Olson, 2005). The understanding of children's mathematics development has evolved significantly since the early work of Piaget. Piaget's work revealed that young children understand far more about mathematics than initially recognized (Starkey & Klein, 2000). Piaget (1952) described mathematics development in a series of stages that necessarily precedes the other (Ojose, 2008), with children more-or-less taking the same path towards mathematics knowledge. However, these stages, as originally proposed, did not account for the variability in children's learning and development trajectories (Vygotsky, 1978).

There are several factors that impact mathematics achievement in kindergarten. These include mathematics and school readiness, kindergarten readiness, number sense, and general cognitive abilities. *"Readiness"* in general describes the prerequisite skills that are necessary for success in a certain domain or experience (mathematics and kindergarten) in this case. *Number sense* is a broad term used to describe a collection of skills and knowledge that promote early number knowledge and ability to manipulate number. Lastly, general cognitive abilities include children's strategy use when solving problems and working memory, which affects the cognitive demand when working through a task—particularly with respect to speed. Mathematics readiness and kindergarten readiness, number sense, and cognitive ability are closely related; together, they influence the trajectory of children's early mathematics development and achievement.

Mathematics and School Readiness

"Mathematics readiness" is a term used to describe the prerequisite skills that are required to succeed in early mathematics classroom. Specifically, Starkey and Klein (2000) outline three key issues surrounding young children's mathematics readiness: child's developing mathematical knowledge, child's ecological context proceeding school entry, and developmentally appropriate expectations with respect to the foundational knowledge and skills necessary to formal school entry in kindergarten. These three issues may serve as a useful framework for understanding the myriad structures that support or obstruct access or prevalence of the experiences that support mathematics readiness for African American children.

The necessity of mathematics readiness is particularly salient for children who are most vulnerable to school failure, as readiness promotes future learning. Welsh, Nix, Blair, Bierman, and Nelson (2010) examined cognitive skills and academic readiness in low-income preschool children. There were several findings of interest in mathematics. Using path modeling, Welsh and colleagues found initial executive function (as measured by word span, peg tapping, change card sort, walk-a-line-slowly, task orientation) predicted growth in numeracy skills in pre-kindergarten; most interesting, this was a reciprocal relationship; initial levels of emergent numeracy skills and the growth in emergent numeracy skills made a unique contribution to kindergarten math achievement. The results of this study serve as evidence of what many researchers and educators have known for many years: initial skill level impacts future achievement.

Several non-experimental studies illustrate the critical nature of school readiness in young children. Approaches to learning (ATL) is a diverse set of characteristics used to measure school readiness, which is described as individual characteristics and behavior that are observed while engaging in learning activities; these include persistence, motivation, initiation, flexibility, and attentiveness (McWayne, Fantuzzo, & McDermott, 2004). An analysis of data from the Early Childhood Longitudinal Study Kindergarten-Cohort (ELCS-K) revealed that beginning in the fall of Kindergarten, each unit increase in early ATL was likened to .38 of an additional point in mathematics (Li-Grining, Votruba-Drzal, Maldonado-Carreño, & Haas, 2010). By the end of the fifth grade, children with better ATL (as defined by 1 SD above the mean) scored .56 SD above the mean than children with less adaptive ATL (defined as 1 SD below the mean) (Li-Grining, C., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K, 2010). Therefore, readiness as measured by approaches to learning may have a long-term impact on growth of mathematics achievement during primary education.

Transition to Kindergarten

The transition to kindergarten presents a unique set of challenges for many early learners. For countless children, kindergarten represents the first formal learning experience (Ray & Smith, 2010). Many educators (and children alike) find this transition to kindergarten rather abrupt with respect to curriculum, teaching style, and classroom structure (Barbarin & Miller, 2009). The difficulty of this transition may be exacerbated by an absence of pre-kindergarten exposure or participation in a more play-based program that places little emphasis on formal instruction. Given the awareness that this transition often proves problematic for long-term success, there is nascent a FirstSchool/First School movement that seeks to address this issue with a pre-k though 3rd (P-3) grade school structure (Barbarin & Miller, 2009). The emphasis of this movement is integrating curriculum and applying appropriate pedagogy to ease the transition to the upper elementary grades (Barbarin & Miller, 2009). Of critical importance is the fact that the FirstSchool movement seeks to provide “safe, healthy, and positive environments, stimulating curricula, and a broad range of instructional practices,” (Ritchie, Maxwell, & Clifford, 2007, p. 87) which often elude African American students, Latino students, and those growing up in poverty.

An important dimension related to the quality and access to pre-kindergarten, and programs that promote readiness for kindergarten, is *learning opportunities.* Learning opportunities are defined as “a set of theoretically driven dimensions of interactions between adults and children with empirically supported links to children’s social, emotional, and academic development” (Hamre & Pianta, 2007, p. 50). While there are several approaches to explain *opportunities to learn*, the most basic factor is the amount of instructional time in the classroom spent on learning. An analysis of the NCEDL MS/SWEEP revealed that on average, children spend approximately 42% of the day *not* engaged in learning activities (Early et.al, 2005; Hamre & Pianta, 2007); this may include transition times, nap times, recess. Mathematics is an area in which opportunity to learn has a direct impact on mathematics achievement; this learning is not incidental, but must be implemented through well-planned deliberate instruction (Ysseldyke, Betts, Thill, & Hannigan, 2004). In addition, many school structures impact the access to quality schooling—such as tracking (see Oakes, 1987 1990; Kilgore, 1991; Carbonaro, 2012; Gamoran & Mare, 1989), and limited access to gifted programs (see Ford, Harris III, Tyson, Trotman, 2002; Smith, LeRose, Clasen, 1991; Ford & Grantham, 2003; Ford, 1989), etc, however, these are less salient or not applicable at the pre-kindergarten level.

Teacher Attitudes

Understanding the practices common to pre-kindergarten and kindergarten programs is inadequate to explain mathematics readiness for African-American and low-income students.

While early screening may readily identify areas of weakness in mathematics, teacher behavior(s) and attitudes in the classroom teachers either promote or hinder present and mathematics achievement. . One issue of major concern is that teachers from higher socioeconomic status (SES) tend to have lower expectations for minority and poor children (Rimm-Kaufman, Pianta, & Cox, 2000). These judgments are complex given they reflect children's actual skills and competencies, expectations upon kindergarten entry, and teachers' personal attributes (such as ethnicity) (Rimm-Kauffman, Pianta, & Cox, 2000). In addition, the ethnic match between teacher and child may affect teachers' subjective evaluation of children (Rimm-Kaufmann, Pianata, & Cox, 2000). The question that naturally follows is the connection between teacher expectation and student achievement outcomes, and whether or not teacher expectations have an enduring impact on young children. Hinnant, O'Brien, & Ghazarian (2009) measured teacher expectations and student achievement at first, third, and fifth grade. At fifth grade, ethnicity and sex informed teachers' expectations of student performance in mathematics. Children classified as minority had lower teacher expectancy scores. At third grade, first grade teachers' expectancy scores had a significant relationship to children's mathematics performance; however, there was a significant interaction with family income. At fifth grade, both first and third grade teachers' expectations predicted math performance. This finding is of particular interest because it reveals that even when teachers overestimate student abilities, with respect to test performance, children are more likely to perform better. The converse is true as well; when teachers have a more negative view of children's mathematics performance, children tend to experience worse performance in future years. Minority boys experienced the lowest performance when their abilities were underestimated (with respect to test scores), and the greatest gains when their scores were overestimated. The authors' findings are not to be taken lightly, particularly with respect to minority males. These findings suggest that children are aware of teachers' perceptions and perform in a related fashion, which is of particular concern to children most vulnerable to negative learning outcomes in education settings.

Teachers' beliefs about race and ethnicity, and their relationship to classroom instruction, impact teachers' beliefs about how to educate African American, Latino, and ethnic minority students. Using the scale adapted by Ladson-Billings (1994), Love and Kruger (2005) surveyed teachers about their culturally relevant teaching practices and beliefs. Teachers from this study, as well as those from prior studies, believed that it was important to incorporate students' culture, race, and ethnicity into classroom teacher practices. Despite teachers' affirmations of the importance of children's backgrounds with respect to instructional practices, teachers still affirmed two statements associated with "color-blindness," agreeing that they did not see race, but rather just saw children. The contradictions in these two perspectives are striking as they suggest some ambivalence in teachers' treatment of race in culture and instruction. It is difficult to envision how teachers could actively incorporate children's backgrounds when they were attempting to ignore these differences.

Teacher preparation programs are culpable in exacerbating the difficulty in educating African American and ethnic minority children in today's increasingly diverse classroom settings. Many of the pre-service field experiences that teachers undergo take place in primarily White middle-class classrooms. As a result, there is significant dissonance between their classroom experiences and the reality of primarily African American urban classrooms (Ladson-Billings, 2000). Many White pre-service teachers have unrealistic expectations about their first placements as teachers. They are unaware that they are most likely to find positions where the need is highest, which is in urban schools, primarily populated by students of color (Swartz, 2003). Predominately White teacher education programs have redoubled this myth and the consequences of this belief about the students there are likely to teach. Most schools do not require pre-service teachers take a multicultural education course, or even fail to offer these courses altogether (Sleeter, 2001). Despite most teachers' beliefs that all children can learn, many are woefully prepared to teach low income, ethnically diverse students—often considered difficult to instruct (Foster, 2004). Urban schools are more likely staffed with teachers that are uncertified or not highly qualified in the subjects that they are assigned to teach (Foster, 2004; Rosas & Campbell, 2010). An additional factor that affects a teacher's ability to succeed with diverse students is the ways in which teachers are assessed with respect to certification and competency to excel in high needs schools. Ladson-Billings (1998) suggests that assessing teachers' cultural competency should be an added factor, as current teacher assessments do not capture the diverse, creative, and adaptive ways that teachers meet the needs of urban students.

Teacher beliefs and attitudes are not the only factors that influence teachers' abilities. Teacher knowledge and instructional practices play a pivotal role as well. Arnold, Fisher, Doctoroff, and Dobbs (2002) conducted an intervention intended on accelerating mathematics development for children in Head Start classrooms. The intervention increased mathematics activity in daily classroom routines, and did so in a way that was flexible for teachers. Children in the treatment group experienced significantly greater improvement on emergent math skills over the children in the control group. Importantly, Puerto Rican and African American children showed greater gains than White children. The authors propose that this might be explained by the variety of group activities that may particularly benefit these groups. In addition, teachers and children that participated in the intervention reported greater interest in children's activities. An alternative explanation, not proffered by the authors, is that the activities and exposure provided through the intervention were distinct enough from those experienced outside the classroom, causing the children to realize substantial growth. In order to address teacher competency for teaching mathematics to young children, interventions need to integrate well into existing classroom routines, and increase teacher competency in mathematics.

Policy and Early Learning

In addition to teacher classroom behaviors, policy has focused on students often considered at greater risk for negative learning outcomes, which is ultimately connected to mathematics achievementLower-income students have frequently been the focus of early intervention programs like Head Start (Reynolds, 1989; Kagan & Kauerz, 2007). Unlike child-care or early education programs, Head Start is intended to meet the educational *and* social needs of children (Kagan & Kauerz, 2007). Given the critical nature of early-intervention and preschool programs, particularly for low-income, and or African American children, the quality of these programs is of particular interest. Extensive work by Pianta and colleagues have explored the vital interaction between positive classroom environment and teacher/child interactions and learning gains (see Burchinal, Peisner, Feinberg, Pianta, & Howes, 2002; Downer & Pianta, 2006; Mashburn & Pianta, 2010). More specifically, there is research on whether there is a minimum quality threshold that is necessary for positive learning and developmental outcomes. Burchinal, Vandergift, Pianta, and Mashburn (2010) examined these relationships extensively and their research revealed that more pre-kindergarten classrooms than not, had higher levels of Emotional Support (as measured by the CLASS), however, fewer of these classrooms were located in Head Start programs, comparatively. Of greater concern is the fact that Black children were more likely to be in classrooms with lower Emotional Support. This finding is consistent with prior work on teacher-child interactions (Huston &McLoyd, & Coll, 1994). Given the known relationship to emotional support and positive learning outcomes, it may be fair to conclude that African American children are at risk for comparatively worse learning outcomes based on classroom environment.

While it is well known that home and school interactions are critical for children's development, the classroom experience is even more important for children who are at greater risk for academic difficulty (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002). Poverty is often considered a risk factor for such difficulty (Brooks-Gunn & Duncan, 1997; Myers, Kim, & Mandala, 2004; Jordan, Kaplan, Olah, & Locuniak, 2006). To further complicate the situation, race and poverty are often challenging to disentangle, which also contend with generally unacknowledged racial and ethnic prejudice (Spencer & Markstrom-Adams, 1990). Despite the fact that most children in the United States are White, African American children and Latino children disproportionately live in poverty (Huston, McLoyd, & Coll, 1994). There is consistent evidence that mathematics performance differs by socioeconomic status, with the typical pattern being that higher income students perform at higher levels than lower income students (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006; Jordan, Levine & Huttenlocher, 1994). However, Jordan, Levine, and Huttenlocher (1994) did not find a significant difference in growth over the preschool year based on SES, suggesting that student outcomes are highly influenced by initial skills development and the quality and diversity of mathematics experiences.

Literacy and Mathematics. Mathematics development is not only impacted by the quality and extent of early experiences related to math, but is connected to literacy-related skills as well. The relationship to vocabulary knowledge and academic achievement has been widely explored in the reading literature (ex. Hart & Risley, 1995; Wright et al., 2001; Verhoeven, van

Leeuwe, & Vermeer, 2011). However, this has been less so in mathematics, although it proves critical. Klibanoff et al. (2006) found that when teacher speech was coded for mathematics related conversation, there was a significant relationship between teacher input in math speech and overall student growth in mathematics knowledge. Surprisingly, the frequency of this speech did not vary by socioeconomic status. However, since lower income students are likely to have lower vocabulary development, it is fair to infer that more than par instruction is essential to improve mathematics gains that may close the gap between lower and higher income students.

Structural and Economic Factors

Student Mobility. Myriad structural and economic factors impact access to quality schooling for low-income and African American children. For example, low-income children are more mobile than those from middle-income or higher-income backgrounds. This mobility is problematic given that evidence that children who remain in the same school for first grade as the school they attended in kindergarten showed greater gains in reading and math than those who changed schools (Reynolds, 1989). Kerbow (1996) studied student mobility in Chicago public schools retrospectively starting in spring 1994, going back two years. Over 36% of students changed schools at least once, 13% of students attended three or more schools, and 5% attended four or more schools—attending multiple schools in one year. The characteristics of mobile students exemplify the intersection of race and class. African American students comprise the largest percentage of mobile students. In addition, mobile students have lower household incomes than their more stable counterparts, and children with mother-father households experience greater stability. Over 46% of stable students live in two-parent households, whereas only 22% of mobile families live in this family structure. Alexander, Entwisle, and Dauber (2001) conducted a similar study in Baltimore over the course of five years, revealing diverse reasons for mobility and differing patterns of movement based on race and socioeconomic status. Children who were more economically advantaged and White tended to move out of the school system to schools that may provide better educational opportunities, and were more likely to be classified as high achievers. Lower income and minority students tend to move within the district, and are more likely to experience academic difficulties. Contrary to the study's hypothesis, there was only weak evidence to support the notion that frequent movement has a significant negative effect on student achievement as measured in this study. However, the authors suggest further research given the family characteristics of children that face these challenges. Regardless of the findings in either study regarding student mobility, it stands to reason that when students change schools, changes in curriculum and school structure are likely, and the discontinuity creates an additional obstacle to achievement—particularly in mathematics—given its naturally hierarchal nature.

Number Sense

There are numerous factors that shape early mathematics development. At the core, however, there are several fundamental skills and competencies that have been identified as critical to this development. Number sense is a collection of skills and competencies that marks early mathematics development and lays the foundation for future success. Number sense is not a singular characteristic or ability, but rather a set of skills that result in mastery of basic number properties and advanced number skills (Aunio & Van Luit, 2005). Berch (2005) describes number sense as "an awareness, intuition, recognition, knowledge, skill, ability, desire, feel, expectation, process, conceptual structure, or mental number line" (p. 1). Gersten and Chard (1999) depict number sense as a child's ability to fluidly and flexibly work with numbers, make numerical comparisons, and perform mental mathematics. In addition, Aunio and Van Luit

(2005) operationalize number sense as the dynamic ability to understand the meaning of numbers and form correct mathematical statements. Number sense has also been defined as the capacity to compare and classify numbers, exhibit one-to-one correspondence, and seriate (Smith, 2002). Kalchman, Moss, and Case (2001) define number sense as being comprised of the following: "(a) fluency in estimating and judging magnitude, (b) ability to recognize unreasonable results, (c) flexibility when mentally computing, (d) ability to move among different representations and to use the most appropriate representation for a given situation, and (e) ability to represent the same number or function in multiple ways, depending on the context and purpose of this representation" (p. 2). While the fundamentals of number sense are generally developed during the primary years of elementary school, deficits in this area are also witnessed and prevalent in students experiencing mathematics difficulties at first and second grade (regardless of race or socioeconomic status). The current study focuses on the predictive validity of number sense in kindergarten children. Below is a review of the variables assessed by these measures.

Most children enter kindergarten with some level of number sense, although it may vary widely (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006; Jordan, Kaplan, Locuniak & Ramineni, 2007; Chard et al., 2008). Other children may require explicit instruction to develop number sense (Chard et al., 2008; Bruer, 1997; Griffin, 1998). It is particularly important to explore the development of students that experience mathematics difficulties as they are more likely to be low-income and/or African American and Latino students given the quality and quantity of mathematics experiences they have prior to kindergarten and the years that follow. Children from high-poverty settings are more likely to experience mathematics deficiencies at the pre-kindergarten level—difficulties that will most likely persist without deliberate intervention (Griffin & Case, 1997; Vanderhyden, Broussard, Fabre, & Stanley, 2004).

Subsequently, knowledge of the challenges experienced by children that later experience difficulties has driven assessments designed to catch children before they begin to strugglePlace value is not directly measured in the present study.

Number recognition/identification. Number identification/recognition has been identified as one of the key number sense skills in several studies and through assessments of number sense (Clements, 1984; Griffin et al., 1994; Payne 1990; Malofeeva, Day, Saco, Young, & Cianco, 2004). Siegler (2009) includes number identification as a key requisite skill in the development of number sense and a precursor to a variety of mathematics skills. There is little discussion of number recognition in isolation in the academic literature, but is regarded both explicitly and implicitly as essential to most mathematical operations. Number identification is a subtest in the TEN requiring students to verbally identify the numbers 0 through 10 in 56 items; each numeral is identified several times within a 1-minute time constraint. The NSB requires students to identify the following numerals: 13, 37, 82, and 124.

Quantity Discrimination. Children's concept of number is critical given its correlation to overall math achievement. Consequently, children's numerical estimation ability serves as a barometer for mathematics development during kindergarten and the years that follow. Outlining this numerical progression demonstrates typical development for kindergarten children and explains the relevance of measuring these skills on early screening measures for kindergarten children.

The mental number line is defined as the system in which numerical quantities are internally encoded on a continuum (Izard & Dehaene, 2008). The mental number line is thought to be a horizontal line in which smaller numbers are on the left and larger numbers are on the right (van Galen & Reitsma, 2008). Number magnitude is the differentiation of numbers based on their relative value (Girelli, Lucangeli, & Butterworth, 2000). The ability to discriminate the magnitude of larger multi-digit numbers is critical as students advance to first and second grade, confronting double digit addition and subtraction problems (Siegler & Booth, 2004). During kindergarten, first, and second grade, numerical estimation on the number-line task correlated with overall math achievement (Booth & Siegler, 2006). Students that represent number linearly solve arithmetic problems with greater accuracy and are more likely to correctly retrieve the correct answer (Booth & Siegler, 2008).

Young children and students experiencing mathematics difficulties possess an informal (non -school-taught) concept of number line in that they have some understanding of the relationship between numbers, such that the child understands that 12 is closer to 10 than it is to 20 (Russell & Ginsburg, 1984). Informal knowledge may stem from everyday experiences with spontaneous counting, using non-mathematical terms. For example, this may occur when children count rocks on the playground or grouping action figures. Students may experience difficulties because of dissonance between this informal knowledge and formal written instruction, which also includes formal mathematical terminology. Even when children begin to develop formal school knowledge, they assimilate, rather than abandon, informal knowledge (Ginsburg & Russell, 1981). Therefore, we would expect that children who fail to successfully assimilate this knowledge and abandon some practices, when appropriate, (or do not receive instruction that does not integrate this knowledge), will experience difficulties.

There is a significant relationship between age and experience with respect to number magnitude estimation. Numerical estimation and mental number line are related to other competencies of early number sense. Therefore, accurately measuring these skills at kindergarten may shed light on the development of other skills such as addition and subtraction ability.

Additional information on the number line/number magnitude is found in Appendix B. Number magnitude is assessed in both the TEN and NSB. There is a number magnitude subtest in the TEN that requires students to identify the larger number in a set of two numbers. The NSB assesses number magnitude in series of seven questions that require students to identify the larger or smaller number in a pair of numbers. In addition, students must identify the number that is closest to a given target number.

Counting. Counting ability is assessed in various forms on early screening measures, as it precedes the development of other early numeracy skills. However, there are varied beliefs on the timing of the inception of counting ability and its role in early numeracy ability, and it remains difficult to ascertain when number understanding begins, as it is unlikely a single entity (Dowker, 2008). Notwithstanding, children's counting ability follows a relatively predictable pattern. Dowker (2005) found that most 4-year olds could count relatively proficiently with sixty-two percent of the sample able to count at least 10 objects correctly. The majority of students (70%) counted objects when prompted by a specific number in the cardinal word principal task, while the remainder of students arbitrarily grabbed the number of objects, underscoring the fact that children develop counting abilities relatively early. Children are expected to accomplish the following at the completion of kindergarten: count to 100 by ones and tens; count forward from a number without starting at 1; write the numeral 0 to 20, and create a corresponding set to represent these numbers (Common Core Standards Mathematics, 2010).

While children develop counting skills early, there is not a seamless transition to mature counting patterns. Gelman and Gallistel (1978) challenged the tendency to assume that children follow the adult pattern of counting, and also countered the notion that counting is achieved when all components of counting are utilized. This alternative perspective has facilitated a greater understanding of the various stages or levels of counting sophistication exhibited by young children. Piagetian tasks (1952) involving conservation were expected to occur in succession; this also includes the stages believed necessary to count. However, children can possess some facets of counting skills while still developing others (Gelman & Gallistel, 1978), debunking earlier notions of an “all or nothing” developmental perspective. Sarnecka and Carey (2008) addressed the knowledge and timing of children’s counting skills and reported myriad findings too numerous for this review, however, it is critical to know that they found that children’s understanding of the “how many” task changes as counting grows, lending support to the progressive development of counting skills. The Common Core Standard K.CC. 5 (2010) requires that kindergarten students answer the question “how many” when presented between 1-20 objects in an array or other organized fashion, and up to 10 when objects are presented in a scattered display, suggesting that this task is reasonable for children of kindergarten age. Nevertheless, counting predicts and precedes the development of other mathematics skills. Fortunately, the development of counting skills is directly connected to opportunities for practice, and the extent of such practice (Gelman & Gallistel, 1978). Fischer, Kongeter, and Hartnegg (2008) conducted an intervention with students experiencing arithmetic difficulties and after 3 weeks of deliberate practice, students in the intervention group improved their counting ability. Children that performed most proficiently on a counting task at the end of kindergarten exhibited better performance on arithmetic tasks in first grade (Stock, Desoete, and Roeyers, 2009). Ultimately, the progression of counting skills in kindergarten can either impede or support the growth of future arithmetic ability for young children. Counting skills were assessed on the TEN Oral Counting measure, which scores the maximum count (up to 100) that a child achieves within a 1-minute period time constraint. Counting was measured by a maximum count question on the NSB; the question was scored correct if the child reached 10; if they were allowed to count higher by the tester, the highest number reached was recorded anecdotally.

Non-verbal calculation (counting objects). Sensitivity to quantity is observed as early as infancy (Wynn, 1992; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Chard et. al, 2008). It is widely documented that touching gestures during counting facilitate and improving counting accuracy (Fuson & Hall, 1983; Gelman & Meck, 1983; Saxe & Kaplan, 1981; Shaeffer, Egleston, & Scott, 1974; Alibali & DiRusso, 1999), suggesting that children become more sophisticated in their approach to counting. Yet some children experience challenges with counting accuracy. Slow counting may be attributable to the phonological representation suffering from short-term memory decay before the count is completed and accessed in long-term memory (Bull & Johnston, 1997). Unlike verbal tasks, children from both lower socioeconomic backgrounds generally perform equally to their wealthier peers on nonverbal numerical operations (Ginsburg & Russell, 1981; Jordan, Huttenlocher, & Levine, 1992; Jordan, Levine, & Huttenlocher, 1994; Ramani & Siegler, 2008). This skill was assessed in four items on the NSB by showing children a target number of dots that is hidden under a box, and then revealed. Children are then shown a single paper with four boxes that different numbers of dots within the box. One of the boxes has the number of dots that was shown under the box; children point to the box that has the correct number of black dots. Nonverbal calculation abilities were assessed in a similar manner by Jordan, Levine, and Huttenlocher (1995).

Word/story problems. Solving story problems can pose a significant challenge for students with mathematics difficulties (Hanich, Jordan, Kaplan, & Dick, 2001). One of the complexities of mathematics problem solving is that it may load heaving on working memory

(Zheng, Swanson, & Marcoulides, 2011). See LeBlanc and Weber-Russell, 1996; Passolunghi and Siegel, 2001; Swanson et al., 2008; Swanson and Sachse-Lee, 2001 for review on working memory and mathematics problem solving. Initially children solve problems using concrete representations/objects to derive the solution strategy (Briars & Larking, 1984; Capenter & Moser, 1984; Fuson, 1988; Riley, Greeno, & Heller, 1983; Willis & Fuson, 1988). Typically, students' strategy use becomes more sophisticated with age (Ilg & Aimes, 1951, but there is limited research on problem solving for children that are at risk for mathematics difficulties (Jitendra & Ping, 1997). Problem solving is especially challenging for many children because it requires competency with language skills and arithmetic skills. Kintsch and Greeno (1985) proposed a framework for understanding the intersection of comprehension and problem solving, which identified three key features: reference to sets, special presuppositions, and comprehension strategies. A particularly relevant assumption is that through instruction on arithmetic word problems, children will understand specific words such as "have, give, all, more and less (p. 111)" in a task specific manner. This is critical in general for the ability to solve a word problem, but is a persistent challenge for students whose first language is not English. On average, English Language Learners perform below their fluent English-speaking peers on standardized mathematics assessments (Abedi & Lord, 2001). In addition, a student's proficiency in English does not necessarily reflect or transfer to mathematics language proficiency (Moschkovich, 1999). English language proficiency is pertinent in the proposed study as a large fraction of participants are Latino—many of which may not speak English at home. Children's problem solving skills are assessed on the NSB; children are orally presented five arithmetic problems.

(Cognitive) Strategy Use. Directly connected to working memory (discussed in the following section) and number sense development is strategy use. Specifically, cognitive strategies are described as conscious, intentional, self-aware cognitive activities (Bjorklund, Hubertz, & Reubens, 2004) utilized to solve mathematics problems. Cognitive strategies are a form of number sense in that they reflect a move away from concrete representations to a more abstract representation and manipulation of number. Children typically make this transition in early elementary school. Children of ages five and six tended to use counting all with a focus on counting concrete objects, but that eight and nine year old children were more likely to use count mentally or retrieval to answer the same problems (Ilg and Ames, 1951). Strategy was scored on the NSB in the Story Problems section and Number Combinations section, which include addition and subtraction problems. The instructions require circling the appropriate strategy however, the list of strategies was not included in the version sent by the author. Therefore, strategies were categorized as the following: cognitive (mental) or concrete representations (pictures/figures), or fingers.

Place value. Place value is a critical part of number sense, although it is not well understood by students (Nataraj & Thomas, 2007). Specifically children fail to understand what the columns ones, tens, and hundreds represent (Varelas & Becker, 2007). For example, a student may view the number 23 as having a value of 2 and separately 3, neglecting the position of the 2 in the tens place as representing a value of 20. Additionally, while children may be able to identify a number with respect to the column in which it is located, they may be unable to discern the relationship between the numbers, particularly when the value of the number changes, altering the values in all the columns. Ward (1979) conducted a study in which he asked 10-year olds to name the number after 06299. Only 41% percent of students were able to complete this task suggesting that even 10-year olds frequently have a poor number sense for place value (as cited in Thompson, 2000).

The literature on children with mathematics difficulties indicates that these children are less likely than typically developing students to understand place value (Hanich, Jordan, Kaplan, & Dick, 2001). This finding has prompted recommendations to introduce place value concepts as soon as students start to work with two-digit numbers, which would begin as early as kindergarten (Common Core Standards, 2010). Kindergarten students should work with numbers 11 to 19 in order to compose and decompose numbers into tens and ones (Common Core Standards, 2010). This may be accomplished through discussions on the nature of two-digit numbers during the primary grades, which may serve as a foundation for strong place value understanding (Baroody, 1990). For students with mathematics difficulties, Baroody (1990) also suggests explicit focus on place value may minimize the reinforcement of numbers as unitary, which may make it more difficult to transition to a multi-digit understanding of number. In an intervention conducted with first and second grade students with mathematics difficulties, Bryant et al. (2008) dedicated additional instructional time to double-digit numbers, and the idea of zero as a place-holder. Both concepts are challenging for children with mathematics difficulties. There was no significant effect of the intervention for first-grade students. Aside from the studies mentioned above, there is a lack of substantive research on place value knowledge in students with mathematics difficulties. Only four studies have measured place value knowledge in written calculation in students with mathematics difficulties (Andersson, 2008). For young children, weaknesses in place value may be connected to deficits that emerge over time.

Attentional resources and working memory. Between kindergarten and second grade, most children shift from a logarithmic pattern of estimation of a linear representation of number when estimating to 100 (Booth and Siegler, 2008). Second grade students produced a generally logarithmic pattern of numerical estimation for quantities between 0 and 1000, and an increasingly linear estimation pattern for quantities between 0 and 100 (Booth & Siegler, 2006). Children that represent number linearly are more likely to produce accurate responses to larger addition problems or near misses when responding incorrectly. In addition, students that represent number linearly are more likely to use retrieval for correct responses (Booth & Siegler, 2008).

Despite the attention given to this relationship, the direct connection between working memory and mathematics is not undisputed (LeFevre, DeStefano, Coleman, & Shanahan, 2005; Raghubar, Barnes, & Hecth, 2010). Notwithstanding, working memory may serve as a key component to understanding differences in development and achievement in young children with respect to mathematics learning (Diverene, Lemaire, & Vandierendock, 2008; Hecht, 2002; LeFevre, DeStefano, Coleman, & Shanahan, 2005; Raghubar, Barnes, & Hecth, 2010). Unfortunately, the majority of research on working memory in young children has been concentrated on students with disabilities, rather than typically achieving students (Imbo & Vandierendonck, 2006). Therefore, the knowledge of working memory in students who may experience mathematics difficulties is limited since these students lack cognitive impairments, but still need of greater instructional support like their peers experiencing disabilities.
A common working memory task is forward digit-span where a participant attempts to recall a string of digits (in order) while performing some other processing task (LeFevre, DeStefano, Coleman, & Shanahan, 2005). There is a marked positive relationship between digit span and age; children become faster as they are able to rehearse information with greater speed (Baddeley, 1992). Passolunghi and Siegel (2001) found that children with mathematics

disabilities struggled with both numerical (and verbal) working memory tasks. These findings suggest working memory deficits may significantly impact mathematics performance and development, and may be significant markers for children who perform below average.

Imbo and Vandierendonck (2007) examined the relationship between strategy use and working memory demands in elementary children. Although this study was focused on older elementary children, the findings may suggest the developmental transition that will lead to mathematics success from lower elementary onward. They found that more efficient strategy selection decreased working memory load, which is critical because elementary children use working memory to solve arithmetic problems. Furthermore, the frequency of retrieval increased with age. Directly relevant to the question of typical development of 4-year-olds and 5-year-olds is that short-term memory can be distinguished between working memory (storage and capacity) in children around age 7, presumably signaling students that have begun to use retrieval for basic math facts and strategy use for unknown solutions. Kindergarten students are expected to fluently add and subtract within 5 (Common Core Standards Mathematics (KO.A.5, 2010), which requires the use of retrieval. Research supports the inference that children who use retrieval most often and most efficiently are more likely to succeed in early mathematics.

CHAPTER 3

Cut off Scores, RTI and ROC Analyses

The research on the usage and development of cut scores in medical practice is diverse and robust. Yet, the use of cut scores in early mathematics screening suffers from several limitations. The majority of research in cut scores has been focused on psychological diagnostic tools and medical screeners. However, the existing research on cut scores in content assessments and for use during the response to intervention (RTI) tiers is problematic at best. Several factors outside the mastery of content frequently influence the application of cut scores in these areas. Given the varied exposure to mathematics instruction prior to kindergarten, the over-identification of students that may be at risk for mathematics difficulties or failure in kindergarten, may result in inflated costs of intervention services for students that may not need remediation (Seethaler & Fuchs, 2010). Based on the inflated costs of special education (Fuchs & Fuchs, 2006), it stands to reason that scores may be artificially influenced by the availability of financial resources in school districts, the access to personnel that are trained to provide remediation, and even lack of precision about the optimal cut scores that will correctly identify students who are likely to experience failure.

The research on the use of cut scores in the diagnosis and prediction of mathematics difficulties is tremendously limited. There is a fair amount of research on developmental differences in mathematics skill development of children that experience poverty; in addition, there is research on developmental differences that may vary by race. However, this research has not extended to the relationship that these factors may have on the functioning of assessments administered to children in these groups. Medical research has gone much further than education research to examine whether or not screening measures are equally effective across racial and ethnic groups—ranging from screening for depression to screening for alcohol abuse (i.e. Mast, Fitzgerald, Stenberg, MacNeill & Lichtenberg, 2001; Aktas, Ozduran, Pothier, Lang, & Lauer; 2004; Frank, Williams, and Bradley, 2008, etc.).

There is no known research on the use of variable cut scores in early mathematics screening. Given the various early learning experiences of children prior to kindergarten entry, it stands to reason that children will have equally diverse instructional needs. Therefore, it may be appropriate to vary cut scores based on children's backgrounds. For example, for a child that has attended a pre-kindergarten program that has formal mathematics instruction, he or she may be best served by having a higher cut score for intervention eligibility. Presumably, their educational experiences will make them less vulnerable to trouble in the future. In contrast, if a student has not had access to pre-kindergarten or has attended a program with lower quality instruction, he or she may merit a lower cut score in order to allow for early intervention, as students from these backgrounds are more likely to experience mathematics difficulties

Response to Intervention (RTI)

Response to intervention is a multilevel prevention system that consists of three levels of action (National Center on Response to Intervention, 2012). Tier 1 of this process is quality core instruction; tier 2 includes moderately intense group intervention targeted at students who need additional support; tier 3 intervention is an intense, individualized instructional program for students that have not responded to tier 2 instruction. Integral to this process is early and ongoing screening and progress monitoring (Fuchs & Fuchs, 2006). Students move throughout these tiers based on data from ongoing assessment. Lack of response to appropriate intervention is typically integral to the identification of learning disabilities (Mastropieri and Scruggs, 2005; Vellutino, Scanlon, Small & Fanuele, 2006; Fuchs & Fuchs, 2006; Hale, Kaufman, Naglieri, and Kavale, 2006). RTI is a process that provides a framework for making data-driven decisions about the appropriate distribution of instructional resources (VanDerHeyden, n.d.). RTI is a proposed method for sustained and systemic improvement in mathematics achievement, which includes two central parts: a process for identifying students that may have learning disabilities and a multi-tiered system of instructional support (Clarke et. al, 2011). Universal screening (for kindergarten mathematics skills in this study) is an integral part of the Response to Intervention (RTI) practice. This form of instructional delivery provides for a more comprehensive diagnosis of disabilities over the prior IQ/achievement discrepancy model (Fletcher et. al, 1998; Pennington, Gilger, Olson, & DeFries, 1992; Vaughn & Fuchs, 2003; Fletcher, Denton, & Francis, 2005) as outlined in the Individuals with Disabilities Education Improvement Act, which places a response to instruction model as a part of diagnosis and eligibility for special education services (IDEIA, 2004).

The use of RTI as an instructional delivery model has been well-documented in the academic literature for reading (see Fuchs & Fuchs, 2006; Gersten and Dimino, 2006; Fuchs, D. Fuchs, L, & Compton, 2004; Simmons, et al., 2008), but less so for mathematics (Clarke et. al, 2011). Nevertheless, it is beneficial for addressing young children's mathematics difficulties (Fuchs et. al, 2006; Fuchs et al., 2008; Hale et al., 2006).There is scant research on RTI in mathematics instruction and the majority of this research is concentrated in Tier II interventions (Clarke et. al, 2011). A database search including PsychInfo and ERIC reveals the absence of response to intervention research in mathematics. The search terms "response to intervention "and "reading" yielded 557 results: "intervention" and "reading" returned 7,371 results. In contrast, “response to intervention” and “math” yielded only 36 results; “intervention” and “math” returned 1,011 results. These results were not screened for relevance or redundancy, however, the vast difference in literacy and mathematics intervention are evident. Seethaler and Fuchs (2005) conducted a meta-analysis across five major peer-reviewed special education journals to determine the proportion of articles about reading and math the evaluated effects of group interventions and the prominence of interventions using randomized control designs. Of the limited numbers of intervention studies published, only 10 of the 44 intervention studies focused on math, and only 4.22% of the 44 intervention studies used randomized control trials (Seethaler & Fuchs, 2005). The lack of published research on mathematics interventions underscores the nascence of response to intervention practices for mathematics instruction in early elementary classrooms.

There is scant research on RTI in mathematics instruction and the majority of this research is concentrated in Tier II interventions (Clarke et. al, 2011). However, there are three proposed central concepts are critical to remediating deficits in number combinations: counting strategies, developing part-whole strategies, decomposition strategies, and improving retrieval from memory (Fuchs et. al, 2010). These skills are related to the PASS cognitive model: Planning includes cognitive control and development of strategies and plans; Attention requires focused cognitive activity and resisting distraction; Simultaneous processing involves the management of several mental processes at one time and clustering data into appropriate groups; Successive processing allows a person to work with information in a specific series or order (Das et al., 1994; Naglieri & Das, 1997; Naglieri & Johnson, 2000). Naglieri and Johnson (2000) applied the PASS cognitive processing theory to an intervention with students receiving special education services in mathematics instruction. Understanding the essential components of Tier II intervention is critical in order to meaningfully respond to student performance in early screening assessments.

Special Considerations in Response to Intervention. A relevant consideration on the effectiveness of response to intervention is culture, race, ethnicity, and language (Linan-Thompson, Vaughn, Prater, & Ciriano, 2006). While a diverse body of literature has emerged regarding the ways that instructional practices and classroom and school environments are, and should be impacted by the demographics of student populations (Alvermann & Xu, 2003; Au & Kawakami, 1994; Squire, MaKinster, Barnett, Luehmann, & Barab, 2003), limited research explores these factors within the context of the RTI model. While RTI should employ empirically supported interventions, there are questions as to for whom, by whom, and in what context intervention proves effective (Cunningham & Fitzgerald, 2006; Klinger & Edwards, 2006). In order to identify a student who has a learning disability, it is critical that he or she receive instruction that that is culturally based and evidence based on the population for whom it is applied (Klinger & Edwards, 2006). Context and diversity are intricately intertwined in the RTI process; students may experience varied success and improvement contingent upon the teacher, setting, and personal traits of the child, thereby making it sometimes difficult to distinguish true disabilities from individual differences in response to instruction or behavior (Donovan & Cross, 2002; Klinger et. al, 2005).

Particularly relevant to the sample in the current study is issues concerning ELLs in the response to intervention/instruction model. Inappropriate assessment has made the under and over representation of children from linguistically diverse groups a critical issue in special education (Donovan & Cross, 2002; Linan-Thompson, Cirino, & Vaughn, 2007). This is particularly so because schools often find it challenging to distinguish between difficulties acquiring English as a second language and actual learning disabilities (Klinger & Harry, 2006; Lesaux, 2006; McCardle, Mele-McCarthy, Cutting, Leos, & D'Emilio, 2005; Wagner, Francis, & Morris, 2005; Rinaldi & Samson, 2008). Although there is much documentation showing the benefits of RTI, little data have been disaggregated to show the effectiveness of intervention practices on ELLs and ethnically diverse students (Donovan & Cross, 2002; Linan-Thompson, Cirino, & Vaughn, 2007). Therefore, for RTI to prove successful with ELL students, learning environments must consider and response to students' cultural experiences along with proven instructional practices (Artiles, 2002; Xu & Drame, 2008).

RTI is not directly applied in this study given that this is not an intervention. However, if these assessments were used by teachers in the classroom settings, the scores could be used to move students through the appropriate tiers based on their response to instruction/intervention. The use of the TEN in this study would most closely correspond to benchmark assessments that are typically administered at predetermined times in the fall, winter, and spring, which are linked to a certain performance standard; performance on a benchmark may be used to predict performance on a high-stakes assessment or qualify a student for intervention services (Ciulla, SoRelle, Kim, Seo, & Bryant, 2011). Ongoing assessment using these measures can indicate student mastery and the need for varying intensity of intervention. The study seeks to validate the NSB and TEN to determine their usefulness in this population as a tool in the RTI process. The current study addresses two primary research questions: (1) which of the early screening measures used in the proposed study offer the best predictive validity? (2)Are universal cut scores appropriate across racial/ethnic, socioeconomic groups?

CHAPTER 4

The Current Study

Several additional studies have stressed the importance of early mathematics screening in kindergarten using different assessments with similar [number sense focused] content, applying similar methods. Lembke and Foegen (2009) assessed students on counting skills and used the TEMA-3 as a criterion measure in first grade. The Test for Early Diagnosis of Mathematical competencies (TEDI-MATH) includes procedural and conceptual counting knowledge and number magnitude comparison, among other skills in kindergarten and was utilized in a study using Kortrijk Arithmetic-Test-Revised (KRT) Arithmetic Number Facts Test (TTR) as criterion measures in the first grade (Stock, Desoete, & Roeyers, 2010). Clarke, Baker, Smollowski, and Chard (2008) assessed number magnitude in kindergarten and used the SAT-10 as a criterion measure in kindergarten. Locuniak and Jordan (2008) assessed word/story problem solving using addition and subtraction in kindergarten and correlated with addition and subtraction tasks in spring of kindergarten. The Number Sets Test requires students identify identical sets (quantity) under a time constraint; this was correlated to WIAT-II as a criterion measure in first through third grade, r=.58 for first grade (Geary, Bailey, & Hoard, 2009). The Number Knowledge Test (NKT) is similar to the NSB in that it is comprehensive—assessing a variety of early number sense skills, including counting, number magnitude, and addition and subtraction skills (Okamoto & Case, 1996; Griffin, 2002). This assessment was administered in kindergarten in with the SAT-9 as a criterion measure in first grade; r=.73 (Baker, Gersten, Flojo, Katz, Chard, & Clarke, 2002). The ASPENS asses the following: a one-minute Number Identification subtest, a two-minute Base-Ten/Number Facts subtest (first grade), one-minute Magnitude Comparison subtest, and a Missing Number subtest (Gersten, Clarke, Dimino, & Rolphus, 2010).

There are several studies that have established the predictive validity the TEN and NSB, a well as various other measures that are not utilized in the proposed study. The TEN was validated by the test developers using the WJ-AP, a Math Curriculum-Based Measures (M-CBM), and the Number Knowledge Test. The predictive validity between fall and spring test administrations of the TEN subtests and the WJ-III Applied Problems (AP) are as follows: Oral Counting (OC), r=.72, Missing Number (MN), r=.72, Quantity Discrimination (QD) r=.79, Number Identification (NI) r=.72 (Clarke & Shinn, 2004). Each of the subtests has strong correlation to the criterion measure, WJ-III AP. However, Baglici, Codding, and Tryon (2010) utilized the TEN in kindergarten and used M-CBM's for predictive validity in first grade, and found only weak to moderate correlations. The NSB is a new measure and has only been cited in one study. It was administered to students in kindergarten and first grade, and then a criterion measure, WJ-III, was administered in third grade. It demonstrated strong predictive validity: r=.63 (Jordan, Glutting, Ramineni & Watkins, 2010). No studies have compared the TEN and NSB screening measures to determine which measure best predicts future mathematics difficulty. The NSB has shown strong predictive validity out to third grade, so it is quite beneficial for predicting mathematics trajectories for young children. However, the administration time is longer than the TEN, which may be a concern for teachers with limited time to administer assessments. Therefore, addressing short-term validity may also inform the way resources are allocated and the benefits of each administration of the assessment.

A variety of skills are assessed through the Test of Early Numeracy (TEN) and the Number Sense Brief (NSB). These include the majority of skills comprise number sense. The TEN and NSB assess counting skills and number sequence, number comparison, number recognition, addition and subtraction skills, story problem solving and strategy use, and non-verbal calculation. These measures, in conjunction with a criterion measure WJ-III AP subtest will provide evidence about the assessments' ability to predict children's mathematical achievement that would be expected from performance on the TEN and NSB. Since both assessments provide suggested cut scores, applying these scores, that are normed on majority White samples, may provide evidence that varied cut scores are needed to most effectively identify students who are likely to experience future mathematics difficulty

CHAPTER 5

Methods

Participants

All participants are kindergarten students from one rural elementary school in school, approximately 100 miles outside of Atlanta, GA. Students were from four kindergarten classes in the school. The majority of students in this school were classified as Latino and African American as is the majority of the sample; this school was classified as Title-I. Title I is a federal program that provides financial assistance to schools with high percentages of children from low-income households with the intention of helping schools reach high academic standards despite poverty status (US Department of Education, 2011). All children were eligible to participate in the study with the exception of students that teachers identified as those who were going to be retained in kindergarten, as they would be ineligible for the follow-up assessment in first grade. Parental consent was obtained for all students participating in the study. The baseline data collection was conducted during March and April, 2012. The sample n=53; the mean age was 5.65 years, 5.92 months, SD=4.53 months. The ethnic/racial composition of the sample is as follows: Black (n=21), Latino (n=28), Mulitracial/Multiethnic (n=3), White (n=1). The final sample for analysis was n=47, including one students that was dropped from the study because of retention, and any additional students that were not tested in first grade because they moved outside the district, changed schools, or were missing a data point. Listwise deletion of subjects (eliminating any subject that has missing data) is how repeated measures ANOVA resolved any missing data point across the two time periods when students were assessed.

Measures

Two different kindergarten mathematics screening measures were utilized in this study. The two measures were the Number Sense Brief (NSB), and the Test of Early Numeracy (TEN). These measures were administered in spring of kindergarten and winter of first grade. Both assessments were administered consecutively in a quiet area of the elementary school.

The TEN has four subtests—each with a one-minute time limit. Oral Counting required students to count as high as they could to 100; Missing Number required children to state the missing number in a string of three numbers; Quantity Discrimination required the identification of the larger number in a pair; Number Identification required students to name the number in a series of numbers.

The Number Sense Brief (NSB) is a comprehensive assessment, with six sections, comprised of 29 items that assess several key early mathematics skills: A) Counting Skills: counting a set of five stars testing one-to-one correspondence and then counting sequencing up to ten; B) Number Recognition: students identified four numbers— from single-digit to three-digits, C) Number Comparisons (magnitude comparisons): select the larger or smaller number in a pair of numbers, D) Nonverbal Calculation: students performed non-verbal calculation problems involving addition and subtraction. Children were presented a number of dots and a quantity of dots was added or subtracted to a group of dots. Students were then required to choose the correct set of dots from four sets of dots in a multiple choice format; E) Story Problems: five orally presented story problems (addition and subtraction); their strategy use was scored on each problem scored; F) Number Combinations: the questions prompted addition and subtraction using the following language: “how much is …and…?” and “how much is …take away…?”

Children's strategy use was noted in this section as well, although it was not factored into the total score in the version of the assessment provided for research purposes.

These assessments were selected for several reasons. The first is their availability—both of these assessments were offered for use for no fee. The TEN is a quick assessment that can be administered throughout the school year with ease, so it is a practical tool for classroom educators, and offers extensive support for schools and systems that subscribe to the AIMSWeb website where the assessments are provided. The NSB also offers relative ease of administration, although it is more time consuming with an administration time of 20-35 minutes per student; however, it has strong predictive validity and can serve as a valuable resource to predict and influence the learning trajectory of kindergarten students. Each of these assessments was administered in the spring of kindergarten. Participants were also assessed in the winter of first grade using the same measures Table 1 describes the characteristics of the early screening measures in detail. In the first grade, participants also completed the Applied Problems subtests of the Woodcock Johnson III (WJ-III) as a criterion measure for the mathematics screening measures. The Applied Problems subtest consists of 39 problems assessing a variety of mathematics skills. The test has various starting points based on grade level. The basal level was established when six consecutive items were answered correctly, and the ceiling was reached when six consecutive items were answered incorrectly, upon which the assessment was terminated.

Table 1

Early Mathematics Assessments and Validity Information

Construct	Author	Measure	Range of Number	Reliability		Validity		
				Sample	Grade	Screening Administered	Criterion	Validity
Counting	*Clarke & Shinn (2004)*	**Test of Early Numeracy (TEN)** Oral Counting (OC) Students count orally in sequence; score is number correctly counted in 1 minute.	0-100	52	1st	Fall & Spring	NKT, WJ-AP, Math Curriculum Based Measure (M-CBM)	M-CBM & OC *r*=.56 WJ-AP & OC *r*=.72 Winter K (OC) & Spring 1st M-CBC *r*=.35
Missing Number	*Baglici, Codding, & Tryon (2010)*	**Test of Early Numeracy (TEN)** Missing Number (MN) Students presented sheet with 21 boxes, each with strings of 3 numbers with 1 missing. Scores is number correct in 1 minute.	0-20	92 K, 61 1st	K, 1st	Winter, Spring	Math Curriculum Based Measure	WJ-AP & MN *r*=.72 Winter K (MN) & Spring 1st M-CBM *r*=.41

Magnitude Comparison	*Clarke & Shinn (2004)*	**Test of Early Numeracy (TEN)** Quantity Discrimination (QD) Identify the greater number from a set of numbers between 1-20; the goal is to maximize number of items completed in 1-minute.	1-20	52	1st	Fall & Spring	Number Knowledge Test, Math Curriculum Based Measure (M-CBM)	QD & M-CBM *r*= .70 QD & WJ-AP *r*=.79
	Seethaler & Fuchs (2010)		1-20	196	1st	Fall & Spring	1st grade Computation and Concepts/Applications (CBM) Key Math-Revised Diagnostic Assessment (KM-R) (Numeration) Early Mathematics Diagnostic Assessment (EMDA) Math Reasoning Test	Spring K and Spring 1st QD and KM-R *r*=.62 QD and EDMA *r*=.47 QD and CBM *r*=.53
	Baglici, Codding, & Tryon (2010)		1-20	92 K, 61 1st	K, 1st	Winter & Spring	Math-CBM	Winter K QD & Spring 1st M-CBM *r*=.22

Number Identification	*Clarke & Shinn (2004)* *Baglici, Codding, & Tryon (2010)*	**Test of Early Numeracy (TEN)** Number Identification (NI) Students orally identify printed numbers between 0 and 20.	0-20	52 92 K, 61 1st	1st K, 1st	Fall & Spring Winter & Spring	NKT, WJ-AP, M-CBM Math CBM (M-CBM)	M-CBM & NI *r*=.60 WJ-AP & NI *r*=.72 Winter K (NI) & Spring 1st (M-CBM) *r*=.41
Comprehensive Assessment	Jordan, Glutting, Ramineni & Watkins (2010)	**Number Sense Brief** (NSB)- 29 items assessing counting, one-to-one correspondence, number recognition, non-verbal addition and subtraction.		204	K, 1st	Fall & Spring	WJ-III (Written calculation and problem solving subtests), Delaware Student Testing Program in Mathematics (DSTP)	Correlation to NSB and WJ-III in 3rd grade: *r*=.63

CHAPTER 6

Data Analysis

Signal Detection Theory

Signal detection theory (SDT) includes the methodof analysis, receiver operating characteristic (ROC) that was utilized in this study. Diagnostic systems, which include ROC, have myriad uses, including the ability to predict future performance; however, it is necessary to assess the accuracy of these systems to determine the best use. (Swets, 1988). Meehl (1959) describes testing as having three general purposes: formal diagnosis, prognosis, and personality assessment. Contemporaries have further categorized diagnostic systems as having two purposes: *diagnostic* assessments are frequently utilized when a person is suspected of having a certain attribute or condition, whereas *screening* tests or assessments were given broadly (Streiner, 2003). Signal Detection Theory is used to analyze data from experiments where it is necessary to distinguish between the known process called the *signal* and the *noise*, which is generated by chance (Abdi, 2007). There was one signal detection methods utilized in this study: receiving operating characteristic (ROC). ROC analysis is useful to support the visualization, organization, and selection of classification criteria (Fawcett, 2005). Receiver operating characteristic curves (ROC) is a preferred method of analysis because it is not influenced by decision biases, prior probabilities, and places classifiers on a common scale (Vivo and Franco, 2008). A *classification model* includes mapping of instances to predicted outcomes, which have four possible outcomes: positive instance classified as a positive is a *true positive* (a participant is *correctly* identified as having a condition); positive instance classified as a negative is a *false negative* (a participant is incorrectly identified as *not* having a condition that they *do* have); if the instance is negative and is classified as negative, it is a *true negative* (a participant is correctly identified as *not* having a condition); if the instance is negative and classified as positive, it is a *false positive* (a participant is *incorrectly* identified as having a condition that is not present) (Fawcett, 2004). This classification matrix is not unlike Type-I and Type-II error. ROC analysis quantifies the extent to which a screener or assessment accurately classified a participant into the correct category based on their condition.

Additional Analyses

In addition to examining the use of universal cut scores and their function across ethnic groups, there was an interest in the overall performance of the TEN and the NSB. Each of these measures was correlated with the WJ-III AP, administered in first grade, to determine which had the best predictive validity. Since the TEN has four subtests, it was possible to determine, which, if any of the subtests was correlated with first grade mathematics achievement (as measured by the WJ-III). In order to explore the stability of the measures across time, the NSB and the TEN were correlated with the WJ-III AP at the kindergarten and first grade time points. Exploratory analyses included ANCOVA, controlling for pretest differences, to determine any differences between the two primary subgroups. However, it was not possible to conduct the ANCOVA due to more independent variables than the SPSS processor could analyze.

Statistical Power

There were numerous challenges in determining the statistical power of an ROC analysis. However, it is critical to outline the basic premise of ROC to understand the challenges experienced in the determination of power in this study. The accuracy (*A*) of an ROC is determined by measuring the proportion or the graph that lies under the curve with values ranging from A=.50 to A=1.0, meaning that a measure has no ability to discriminate to perfect discrimination ability, respectively (Swets, 1988). This area is generally described as the AUC, area under the curve (Henderson, 1993; Jordan et al., 2010). The AUC represents the probability that a score drawn at random from one population (i.e. students not at risk for math difficulty) (on a continuous or ordinal variable) is higher than a score drawn at random from another sample or population (i.e. students at risk for math difficulty) (Rice and Harris, 2005). An important consideration is the computation of sample size based on the parameters of the relevant clinical application (Obuchowski, 2000). ROC performance may vary when the ROC is applied in different situations (i.e. different population) (Fan, Upadhye, & Worster, 2006). Many sample size tables are derived from medical models that employ multiple observers and assumed patient variability, etc. utilizing complex mathematical formulas (Obuchowski, 2000), that are impractical and inappropriate for the use of screening assessments in educational settings. Given the theoretical challenges of determining sample size and power in ROC analysis, simulation is a common method for making these determinations (Rotello, Masson, & Verde, 2008). However, there is no clear method for determining the sample size needed for a specific level of statistical accuracy (Hanley & McNeil, 1982).

Notwithstanding, statistical software packages allow for the determination of statistical power in ROC analysis. There are several statistics necessary to calculate the power of a given sample: AUC, number of individuals with the condition and the number of individuals without the condition, etc. The NSB does provide the sensitivity and specificity statistics at various time points, as well as the AUC. However, there is not specific information about the sample size at various time points, nor the number of children that would be classified as not meeting standards, or failing to meet a certain cut score at a given time point (in this case April of kindergarten).

Several attempts were made to conduct this power analysis in PASS software, which has the capability to provide power analysis for ROC; however, it would require one to make inferences and extrapolations that are not supported based on the information provided by Jordan and colleagues (2008).

Cut scores

The cut score is the point on a test score scale where the examinee is said to have met the performance standard of the assessment (Kane, 2005). However, there are several challenges to evaluating the effectiveness of a given diagnostic tool given the scant reporting methods in the academic literature. These challenges include: failure to report base rates (experience tables of calculated expectations), inability to evaluate the efficiency of the instrument or device across settings, which is also plagued by small sample sizes; there is an absence of cross validation coupled with small criterion groups; there is ambiguity in the population demographics where a device or instrument in appropriate for use; data are reported as significance tests between groups rather than the number of correct decisions by group (Meehl & Rosen, 1955).

The judgments that can be made from an assessment are fundamental to an assessment's validity. However, validity is not an intrinsic property of a test, but rather a property of the interpretations assigned to test scores (Kane, 1994). This study centers on the comparative predictive validity of the two screening assessments being used in the study. Therefore, the primary validity concern for validity in this study is the interpretations that can be made with respect to the population that is assessed and the context in which the assessment is used (Kane, 1994). Although there is substantial research on methods of cut score development with respect to criterion referenced tests, the same rigor and research does not seem to be present, yet, in early screening for mathematics assessments. In addition, these cut scores are universal, rather than specifically adjusted for the unique characteristics of the populations in which they may be used.

There may be several issues that affect the validity of cut scores used to make instructional decisions from early screening measures in early kindergarten mathematics. One criticism of cut scores is what Glass (1978) describes at the "arbitrariness" of passing scores (Kane, 1994). He asserts that it is not possible to determine mastery levels (i.e. criterion levels or standards) by statistical or psychological means, but rather they are done so arbitrarily (Glass, 1978). While this perspective of arbitrariness has been highly contested (see Popham, 1978; Block, 1978, etc.), it may prove valid regarding the determination and use of cut scores regarding early mathematics screening measures. There is limited research on early mathematics screening measures, particularly with respect to the scores that are applied to determine whether or not further intervention is necessary. Much has been reported about the challenges facing schools in urban settings and those that serve lower-income students, and there is some research about the challenges to the RTI process in these schools (see RTI Action Network), however, this work focuses primarily on the global challenges faced by urban and low income schools. Yet, the challenges that may be most significant in the RTI process regarding cut scores, may be the economic and human capital resources available to meet the vast number of students that may experience early difficulties in mathematics. The increasing intensity required as students move to higher levels of the RTI process necessarily involves quality of time, effort, and resources that make intense intervention difficult in traditional school or classroom environments (Barnett, Daly III, Jones, and Lentz, Jr., 2004). As resources are increasingly limited, the cut scores that may be used to make such decisions (at the school or district level) may be artificially adjusted because of the inability to serve all students that may require Tier 2 and Tier 3 intervention support. The cut scores applied in early screening assessments do not undergo the rigor expected and applied to standardized assessments such as the state testing programs. Therefore, these scores merit further research as there is insufficient research to conclude that universal cut scores are appropriate and valid.

There is limited guidance on the development and use of cut scores outside of norm-referenced assessments and criterion-references assessments. Suggestions for related cut score development studies for educational assessments involve expert panels that engage in an iterative process determining the cut scores using the best available information about the population and the assessment (Zieky, Perie, & Livingston, 2008). However, the demands of cut score studies are not feasible given time and resource constraints of this study. Therefore, the cut scores that were validated in the current study were be based on the recommendations of the assessment developers.

The Test of Early Numeracy (TEN) provides national norms and percentile rankings at the various time points (fall, winter, and spring) when the assessment is typically administered during the school year; in a school setting these would be considered benchmark assessments. In addition there are target scores presented for each measure or subtest of the TEN. The suggested target scores (raw scores) for the spring test administrations are as follows: 70 for Oral Counting; 55 for Number Identification; 25 for Quantity Discrimination; 13 for Missing Number (AIMSweb, 2012). The Number Sense Brief (NSB) reports optimal ROC scores and related statistics. The relevant score for the baseline assessment that occurred in April and May corresponds to the kindergarten optimal *d-based ROC* for April of kindergarten: cut score 20; sensitivity .74, and specificity .75, and optimal sensitivity-based ROC with a cut score of 17, and sensitivity .87 and specificity .50 (Jordan, Glutting, Ramineni, and Watkins, 2010). Therefore, the same cut scores were utilized in the sample from this study in order to validate the sensitivity and specificity values found by Jordan and colleagues.

CHAPTER 7

Results

The current study addresses two primary research questions: (1) which of the early screening measures, the NSB or the TEN, demonstrates the best predictive validity? (2)Are universal cut scores appropriate across racial/ethnic, socioeconomic groups? Descriptive statistics are reported in two methods in Table 2: raw scores and conversion scores. Raw scores are reported for all variables. Conversion scores were created for Quantity Discrimination, Missing Number, and Number Identification because the scales changed across grades. The conversions scores, presented in the second column, represent the proportion of correct responses across kindergarten and first grade.

Table 2

Descriptive Statistics

Measures *n*=47	Raw Mean (*SD*)	Converted Mean (*SD*)*
Quantity Discrimination (K)*	21.60 (7.246)	.771 (.259)*
Quantity Discrimination (1)*	30.43 (7.500)	.761 (.188)*
Number Identification (K)*	51.57 (7.174)	.921 (.128)*
Number Identification (1)*	61.62 (14.741)	.770 (.185)*
Missing Number (K)*	12.81 (5.148)	.610 (.245)*
Missing Number (1)*	18.23 (4.833)	.608 (.161)*
NSB (K)	21.47 (4.804)	
NSB (1)	26.00 (2.579)	
Oral Counting (K)	59.70 (13.778)	
Oral Counting (1)	76.23 (13.255)	

* *Indicates score was converted to proportion of correct responses, and the SD reflects the converted variable.*

Predictive validity was determined using two methods. Zero-order correlations are displayed in Table 3 and partial correlations (controlling for language status) are displayed in Table 4. Partial correlations were reported because there was an interest in whether or not language or race significantly affected test performance, so there was an interest in reporting both as a means of controlling for any effect of language on the relationship between variables. Adding the covariate had no effect on the statistical significance of any variable. Correlations only varied slightly when the covariate was added. Correlations demonstrate which subtests of the TEN and the NSB have the highest correlations with the WJ-III Applied Problems, which was administered only once as a criterion measure in the first grade. Several subtests of the TEN were significantly correlated with the WJ-III AP. At the kindergarten time point, Quantity Discrimination r=.411, Number Identification r=.382, and Oral Counting r=.513 were significant at p <.01 level. The NSB demonstrated the highest correlation with the WJ III, r=.629, p <.01. At first grade, the significant predictors of performance on the WJ III AP were primarily the same: Quantity Discrimination r=.483, Number Identification r=.293, Missing Number r=.371, NSB r=.625, Oral Counting r=.311. However, the strength of the correlations changed with the exception of NSB, which remained stable at both kindergarten and first grade.

Table 3

Correlations

Measures	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. WJ-III AP (1)	__										
2. Quantity Discrimination (K)	.411**	__									
3. Quantity Discrimination (1)	.483**	.434**	__								
4. Number Identification (K)	.382**	.645**	.554**	__							
5. Number Identification (1)	.293*	.270	.732**	.523**	__						
6. Missing Number (K)	.282	.353*	.476**	.358**	.440**	__					
7. Missing Number (1)	.371*	.236	.578**	.315*	.460**	.411**	__				
8. NSB (K)	.629**	.588**	.524**	.435**	.413**	.443**	.232	__			
9. NSB (1)	.625**	.341*	.576**	.541**	.530**	.356*	.428**	.530**	__		
10. Oral Counting (K)	.513**	.392**	.267	.366**	.391**	.485**	.101	.456**	.448**	__	
11. Oral Counting (1)	.311*	.275	.127	.226	.094	.161	-.252	.406**	.120	.357*	––

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

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

(K)= Administered in kindergarten; (1)= Administered in first grade.

Table 4

Partial Correlations (Covariate: Language Status)

Measures	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. WJ III	___										
2. Quantity Discrimination (K)	.404**	___									
3. Quantity Discrimination (1)	.474**	.429**	___								
4. Number Identification (K)	.367*	.539**	.544**	___							
5. Number Identification (1)	.274	.260	.713**	.506**	___						
6. Missing Number (K)	.272	.316*	.458**	.234	.418**	___					
7. Missing Number (1)	.354*	.214	.558**	.293	.434**	.400**	___				
8. NSB (K)	.629**	.527**	.504**	.325*	.389**	.442**	.215	___			
9. NSB (1)	.624**	.344*	.562**	.538**	.514**	.339*	.423**	.513**	___		
10. Oral Counting (K)	.526**	.378*	.235	.342*	.374*	.508**	.098	.459**	.426**	___	
11. Oral Counting (1)	.331*	.316*	.131	.235	.083	.153	-.245	.408**	.099	.320*	____

** Correlation is significant at 0.01 level

* Correlation is significant at 0.05 level

(K)= Administered in kindergarten; (1)= Administered in first grade.

A repeated measures ANOVA was conducted to determine the change in student performance on the TEN and the NSB between kindergarten and first grade. This data is detailed in Table 5. There are several assumptions of repeated measures analysis of variance: 1) the dependent variable is continuous; 2) the independent variable must have at least two categorical groups; 3) there should be no significant outliers in the two related groups, 4) the distribution of the differences in the dependent variable and the related groups should be approximately normal; 5) the variance in the differences of possible related groups should be equal. Mauchly's test reveals sample violates the assumption of sphericity, (χ^2=0, df=0, p=.000). Degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε=1).

There was a significant effect of time on student performance on the following assessments: Number Identification $F(1,47)=23.459$, $p<.000$, Oral Counting $F(1,47)=36.193$,

$p<.000$ and the Number Sense Brief $F(1,47)=34.393$, $p<.000$. There was no significant effect of Time*Language status for any assessment. It is important to note that of the assessments where there was a significant effect of time, the Oral Counting measure and the Number Sense Brief assessments tested the same content during kindergarten and first grade, therefore students simply had to answer more questions correctly to demonstrate improvement. However, Number Identification assessed additional content at first grade, in both the number of items assessed, and the difficulty of assessment, so students had a higher standard to meet in order to experience improvement. There was no significant effect of time on the Quantity Discrimination and Missing Number tasks—which required students to answer more items correctly, and have mastery of more content to perform equally well or experience improvement. The implications for these findings are addressed in detail in the discussion. Table 6 details the between subjects effects of the repeated measures ANOVA; all of the values of the intercepts are significant for each variable.

Table 5

Repeated Measures of Analysis of Variance (Within Subjects)

Within Subjects Effects	Measure	*F*	η^2
Time	Quantity Discrimination	.075	.002
	Number Identification*	23.459	.343
	Missing Number	.002	.000
	NSB*	34.393	.433
	Oral Counting*	36.1933	.466
Time*Language Status	Quantity Discrimination	.007	.000
	Number Identification	.010	.000
	Missing Number	.020	.000
	NSB	.006	.000
	Oral Counting	.313	.007

*Significant, $p<.05$

Table 6

Repeated Measures of Analysis of Variance (Between Subjects)

Between-Subjects Effects Measures	*F*	η^2
Intercept		
Quantity Discrimination	462.69*	.911
Number Identification	1103.594*	.961
Missing Number	360.995*	.889
Oral Counting	1439.044*	.970
Number Sense Brief	1005.814*	.957

*Significant, $p<.05$.

The methods for the ROC cut score study of the NSB were derived primarily from the validation study performed by Jordan and colleagues (2010). However, in the absence of a standardized state assessment with established criterion (met/did not meet) standards, several cut points were tested using the WJ-III standard scores. Results of the cut score study are reported below in Table 7.

Table 7

ROC Validated Against NSB Recommended Cut Score for 1st Grade

WJ-III Grade Equivalent (Cut Score)	Students Identified As Positive (Negative)	Sensitivity	1-Specificity (False positive rate)	AUC	Statistical Significance of AUC *p*
1.2	2 (45)	.500	.089	.811	.140
1.4	4 (43)	.500	.026	.878	.013*
1.6	8 (39)	.500	.070	.872	.001*
1.8	14 (33)	.500	.089	.746	.008*

*Statistically significant, p<.01

Table 8

ROC Validated Against NSB Recommended Cut Score for Kindergarten

WJ-III Grade Equivalent (Cut Score)	Students Identified As Positive (Negative)	Sensitivity	1-Specificity (False positive rate)	AUC	Statistical Significance of AUC *p*
1.2	2 (45)	.500	.089	.756	.225
1.4	4 (43)	.500	.070	.735	.123
1.6	8 (39)	.625	.026	.862	.001*
1.8	14 (33)	.571	.030	.820	.001*

*Statistically significant, p<.01

The values reported in Table 7 reflect the sensitivity and 1-Specificity (false positive rate) as reported by the ROC analysis conducted in SPSS. In addition, the table identifies students that would be identified as not meeting the cut score associated with each grade standard score. The raw scores associated with the WJ-III standard scores were dichotomized at each level. Given that the criterion measure used in this study was not the same as that used in the original NSB study, it was essential to show the ROC results and each cut point to demonstrate which cut point proved most beneficial. Given that the WJ-III was administered in spring of first grade, it would most closely be associated with the standard score of 1.6 or 1.8. Therefore it is not surprising that when the higher standard score was tested as the cut point, more students were identified as "positive" for being at risk based on failing to meet to achieve the associated raw score. The AUC scores are moderate to high at each cut point. However, not all of the scores are statistically significant. The AUC values are statistically significant at the following standard score cut points: grade 1.4, grade 1.6, and grade 1.8. However, the high AUC values would suggest an expectation of higher sensitivity scores, although this is not the case. Given that the AUC scores are relatively high, it seems fair to infer that there these scores are not particularly useful for interpreting the diagnostic accuracy of the NSB, but rather the full range of sensitivities found in the ROC curve will provide greater information (Zweig and Campbell, 1993). Possible explanations for this seemingly contradictory finding are addressed in the discussion.

CHAPTER 8

Discussion

The first research question addresses the question of comparative predictive validity: Does the TEN or the NSB demonstrate the best predictive validity? There is not a single response to this question, but rather depends on the goal of a teacher or other stakeholder in the performance on the assessment. Given that these students were in kindergarten and first grade, there is not robust standardized assessment data as criterion reference. While the state of Georgia mandates a kindergarten assessment program called G-KIDS, it would not be useful for addressing the question of predictive validity, as it is isolated to kindergarten. Students to not take begin to take the mandated Criterion Referenced Competency Test (CRCT) Subsequently, the WJ-III was used as this measure.Although it is an achievement test, rather than a criterion-referenced test,which means it does not directly assess standards in which students are expected to receive instruction, it was a common criterion assessment in studies with a variety of screening assessments. If the measure of predictive validity (predicting first grade achievement based on kindergarten assessment data) is the correlation to the WJ-III, the NSB proves the best measure for predicting math achievement, r=.629. This finding is consistent with high predictive validity found by Jordan, Glutting, Ramineni, and Watkins (2010); students with low NSB scores were helpful in predicting student that would fail to meet standards in the Delaware State Testing Program in Mathematics in third grade, which was highly correlated with the WJ-III Mathematics. The NSB assesses a variety of mathematics skills as does the WJ-III. Although the NSB is a mathematics screener, the breadth of skills assessed on the NSB proves similar to an achievement test, and the high correlations with the WJ-III AP are likely indicative of that alignment.

In contrast, the TEN is an assessment comprised of four subtests, assessing key components of number sense: Oral Counting, Number Identification, Missing Number, and Quantity Discrimination. Most of the subtests of the TEN are significantly correlated with the WJ-III AP, although they are not as strongly correlated as the NSB. It is particularly interesting to examine the change in correlations of the subtests to the TEN between kindergarten and first grade. At kindergarten, three of the four subtests of the TEN were significantly correlated to the WJ-III AP: Quantity Discrimination r=.404, Number Identification r=.367, and Oral Counting r=.526. However in first grade, Missing Number r=.354 was significant, and the strength of the correlations changed for subtests that remained significant at first grade: Quantity Discrimination r=.474, increased; Oral Counting decreased to r=.331, although still significant. Interestingly, Number Identification was not significantly correlated to WJ-III at first grade. However, Missing Number was significant, r=.354 at first grade, although it was not significantly correlated at kindergarten. The changes in the values of these correlations may have several important implications. First, not all mathematics skills equally predict math achievement, and the importance of those skills changes over time. Arguably, the most basic skills were oral counting and number identification. Oral counting remained a significant predictor of mathematics achievement between kindergarten and first grade, but decreased in magnitude, suggesting that it may diminish as predictor as mathematics concepts and tasks increase in complexity. Number identification, in contrast, was not a significant correlate to math achievement in first grade. The reason for this relationship was not immediately clear, but it may signal that while children are capable of identifying numbers, the ability to do so may not sufficiently support other mathematics skills. In contrast, Missing Number was a significant correlate of math achievement at first grade, r=.354 although it was not in kindergarten, suggesting the underlying concepts associated with this task are increasingly important to mathematics development and achievement. The ability to discriminate between quantity was increasingly correlated to math achievement in this study, which is congruent with existing research on this skill (Gersten, Jordan, & Flojo, 2005; Geary, Bailey, & Hoard, 2009; Clarke & Shinn, 2004).

The repeated measures ANOVA also provided insight into the stability and growth of certain math skills. The NSB and Oral Counting measures assessed the same content across kindergarten and first grade. Therefore, it was not surprising that students experienced significant improvement on these measures. In contrast, students declined in their ability to identify numbers; at kindergarten, students correctly identified 92% of numbers (1-10). However, when the numbers 1-20 were assessed, students only identified 77% of the numbers correctly. Students experienced difficulty identifying double-digit numbers. In contrast, students were stable in the Quantity Discrimination task and the Missing Number task. The lack of change across grade levels may reveal that students do not receive as much instruction related to these tasks, or that these tasks are more challenging for these students. While students may naturally become better counters, they may need more targeted instruction or intervention to realize improvement in Number Identification, Quantity Discrimination, and Missing Number. Furthermore, since students faltered in their ability to identify the numbers 1-20, they were also unable to perform other tasks with these numbers (i.e. filling in the missing number or discriminating between quantities). This finding is somewhat in line with that of Clarke and Shinn (2004), whose validation study found that Oral Counting was the most sensitive of the measures, followed by Number Identification, Quantity Discrimination, and Missing Number. However, while the Oral

Counting measure showed the greatest improvement, it had the lowest reliability and validity coefficient correlations (Clarke & Shinn, 2004), which is indeed confirmed in this study as well. This finding is not to be taken lightly given that, with the exception of the NSB, Quantity Discrimination has the highest significant correlation to math achievement (as measured with the WJ-III). However, the administration of the TEN occurred slightly earlier in first grade than it did in kindergarten, so it is unclear whether that impacted growth on these tasks.

An important distinction between the subtests of the TEN and the WJ-III AP and NSB is that the TEN is a timed assessment—each subtest was constrained to1-minute in duration. In contrast, there was no time constraint on the WJ-III AP and NSB. Fluency and automaticity may be connected to the performance on the NSB—both of which are speed related. A participant's score was based on the number of correct responses, which is also connected to how many items are attempted. For students who are accurate on this task, but do not have the speed to respond to all questions, they will invariably receive lower scores. Fluency is an integral part of math achievement; the absence of such fluidity with number may signal mathematics difficulties (Gersten & Chard, 1999; Gersten, Jordan, & Flojo, 2005).

The cut score study validating the NSB raises several questions about the conditions under which recommended cut scores are replicable across variable geographic areas, ethnic/racial groups, and economic backgrounds. These concerns are addressed by Jordan et al. (2010), cautioning that the findings should be viewed as preliminary until they are shown replicable across diverse samples. The criteria for being identified as "at- risk" for mathematics difficulties in this study was dependent upon performance on the WJ-III, whose validity is well-established, but may not be ideal for this type of analysis. Furthermore, given the economic and racial homogeneity of this sample, scores were clustered closely together, limiting potential variability in performance outcomes. In addition, the sensitivity scores found in Jordan et al.'s study did conform to the suggested cut scores. It is unclear whether this is a consequence of the different criterion measure used, the slight variation in the timing of the assessments, or simply a characteristic of the sample (i.e. student's skill levels). While there was an interest in whether or not being classified as an English Language Learner would impact mathematics performance, preliminary analyses revealed that there was no significant difference by language background. Therefore, data was not further disaggregated based on language background. It is unknown whether this finding would replicate across samples from different region, sample, or socioeconomic background.

Limitations

There are several important limitations in this study. The first is that students were sampled from a single school, which exacerbates the homogeneity in the sample. The school was classified as Title I, resulting in limited economic heterogeneity. In addition, the sample was primarily Black and Latino. It is possible to argue that more samples should reflect these demographics if they are reflective of the changing demographics of schools. However, it is difficult to know whether the findings in this study were contingent on this sample or if they would replicate in samples of similar demographic characteristics. Furthermore, despite the economic status of the participants in this study, they reside in a county where the overwhelming majority of students have access to quality prekindergarten programs, resulting is scores that are likely inflated compared to students that may not have such access nationally.

An additional limitation is the absence of a valid and reliable standards-based assessment to classify students as meeting standards or not meeting standards. The availability of a state assessment with well-established validity data would provide a deeper understanding of how state standards are connected to a screening assessment that is intended to address critical early mathematics skills. While the WJ-III is a well-validated achievement test, it is not well-aligned to grade-appropriate standards. Subsequently, scores may have been more closely clustered together, thereby limiting score variability. This lack of variability likely impacted the specificity associated with the recommended cut scores for the timing of the NSB. Furthermore, when ROC analysis is used in medical settings, there is typically a “more perfect” dependent measure against which the screener is evaluated (i.e. the outcome of developing cancer or remaining cancer free). Contrary to medical applications, the criteria for evaluating the presence of a condition, (i.e. math difficulty), is more likely to have validity concerns. For example, if the dependent measure is an assessment from a state testing program, the failure or pass scores may depend on a variety of factors extraneous to the mastery of the majority of content standards. For example, a state may have an interest in setting the criteria for passing the assessment at a level to ensure that the majority of students meet the standards.

CHAPTER 9

Implications and Recommendations

Given that assessment is increasingly used as a metric for teacher evaluation and measuring student outcomes, it is necessary to expand the uses and benefits of assessments. If schools were to implement comprehensive screening programs in mathematics as they often do in reading (i.e. IES, 2011; Molfese, Modglin, Walker, & Neamon, 2004; Ardoin et al., 2004; Fuchs, 2004; Elliott, Huai, & Roach, 2007, etc.) teachers would have a more robust data to identify students who may need intervention or modified instruction. In addition, teachers can assess the effectiveness of their instructional practices(s) based on student performance on these screening measures. In addition, districts have the opportunity to provide targeted professional development to support teachers' ability to provide quality, targeted instruction.

Furthermore, in order for assessments to be considered valid for the full range of students that are evaluated using such measures, it is critical to examine whether the inferences made as a result of these assessments are generalizable across regions, ethnic/racial groups, and economic backgrounds. While there was not variability in performance by language status, and the small sample size made disaggregating the data implausible, there is still uncertainty as to whether a single cut scores yields the same quality of inferences across varied conditions and backgrounds. There remains little to no research on this topic, specifically with the respect to mathematics screening assessments. The utility of ROC as a means of evaluating diagnostic accuracy in mathematics assessments is promising, but requires much research to expand our understanding of this tool for this specific application. Furthermore, given the budgetary and time constraints facing school districts, it is imperative that schools allocate resources based on empirically-based data. Lastly, the findings in this study suggest that there are key skills that disproportionately predict math achievement, and such knowledge has the power to alter the impact of instructional time. Policy has often favored a “more is better” approach to improving academic performance (i.e. the disproportionate instructional time spent on reading in federally-funded programs such as Early Reading First) (IES, 2011), which is not sufficient or proven to increase achievement. Instead, screening assessments have the potential to focus the use of instructional time to target skills that are most impactful for mathematics trajectories, rather than blindly increasing instructional time as a means for improving achievement.

CHAPTER 10

Conclusion

Given the inferior mathematics performance of many U.S. students, there is increased interest is preventing mathematics failure instead of relying on intervention. For this reason, mathematics assessments for young children, specifically kindergarten children, are growing in popularity and diversity. The early mathematics screening movement seeks to identify students' strengths and weakness by mathematics domain in order to better ascertain student ability and subsequently improve mathematics instruction. The importance of early screening goes beyond the obvious desire to prevent difficulties. Given the myriad obstacles to mathematics success for the most vulnerable children, early screening may serve to increase equity and access to quality mathematics instruction for students from typically underserved backgrounds. If educators are successful at circumventing the mathematics difficulties that often endure into future years, then children from these groups may see improved educational trajectories. Since quality early learning and kindergarten learning experiences are so integral to future success, comprehensive and ongoing screening serves a vital role in improving long-term mathematics outcomes for all children, and even more so for children that typically encounter schooling experiences that may prove inadequate for long-term success.

References

Abdi, H. (2007). Signal detection theory (SDT). *Encyclopedia of Measurement and Statistics Thousand Oaks, CA: Sage,* 866-88.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education,14*, 219–234.

Alexander, K., & Entwisle, D. (1988). ACHIEVEMENT IN THE FIRST 2 YEARS OF SCHOOL: PATTERNS AND PROCESSES. *Monographs of the Society for Research in Child Development*, *53*(2), 1-140. doi:10.1111/1540-5834.ep11865354.

Alibali, M. W., & DiRusso, A. A. (1999). The function of gestures in learning to count: More than keeping track. *Cognitive Development, 14*, 37–56.

Alvermann, D. E., & Xu, S. H. (2003). Children's Everyday Literacies: Intersections of Popular Culture and Language Arts Instruction. *Language Arts*, *81*(2), 145-54.

Amit, M., & Fried, M. N. (2002). High-stakes assessment as a tool for promoting mathematical literacy and the democratization of mathematics education. *Journal of Mathematical Behavior 21*, 499−514.

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L. & Williams, K. L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review*, *33*, 218-233.

Au, K. H., & Kawakami, A. J. (1994). Cultural congruence in instruction. *Teaching diverse populations: Formulating a knowledge base*, *24*.

Augustine, N. (Chair) (2005) Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future, National Academy of Science, National Academy of Engineering, Institute of Medicine, National Academy Press, Washington, DC.

Aunio, P., Hautamäki, J., & Van Luit, J. E. H. (2005). Mathematical thinking intervention programmes for preschool children with normal and low number sense. *European Journal of Special Needs Education, 20*(2), 131-146.

Baddeley, A., & Larsen, J. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. Quarterly Journal Of Experimental Psychology (2006), 60(4), 497-504.

Baglici, S., Codding, R., & Tryon, G. (2010). Extending the Research on the Tests of Early Numeracy: Longitudinal Analyses Over Two School Years. *Assessment for Effective Intervention*, *35*(2), 89-102.

Baker, S., Gersten, R., Flojo, J., Katz, R., Chard, D., & Clarke, B. (2002). *Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays* (No. 0305). Eugene, OR: Pacific Institutes for Research.

Ball, D.DL. Lubienski, S., Mewborn, D. (2001). Research on teaching mathematics: The unsolved problems of teachers’ mathematical knowledge. In V. Richardson (Ed.), Handbook of research on teaching (4th ed.). New York: Macmillan.

Barbarin & Miller. (2009). Developmental Science and Early Education: An Introduction. In O. Barbarin & B. Lasik (Eds.), *In Handbook of Child Development & Early Education* (pp. 3-13). New York, US: The Guilford Press.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention. *The Journal of Special Education, 38*(2), 66.

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

Bjorklund, D. F., Hubertz, M. J., & Reubens, A. C. (2004). Young children's arithmetic strategies in social context: How parents contribute to children's strategy development while playing games. *International Journal of Behavioral Development, 28*(4), 347-357.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189-201. doi: 10.1037/0012-1649.41.6.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016-1031. doi: 10.1111/j.1467-8624.2008.01173.x

Branscum, A, Johnson, W., & Gardner, I. (2007). Sample Size Calculations for Studies Designed to Evaluate Diagnostic Test Accuracy. *Journal of Agricultural, Biological, and Environmental Statistics, 12*(1), 112-117.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skills in solving elementary word problems. *Cognition and Instruction*, 1, 245– 296.

Brooks-Gunn,J.,Klebanov. P. K.,Smith.J. R., Duncan, G. & Lee. K. (2003). The black-white test score gap: The contribution of test, poverty and the home environment. *Applied Developmental Science, 7*, 239-252.

Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, 26(8), 4-16.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology, 65*(1), 1-24.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, 179– 202.

Chard, D. J., Baker, S. K., Clarke, B., Jungjohan, K., Davis, K., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly, 31*(1), 11-20.

Ciullo, S., SoRelle, D., Kim, S. A., Seo, Y., & Bryant, B. R. (2011). Monitoring Student Response to Mathematics Intervention: Using Data to Inform Tier 3 Intervention. *Intervention In School & Clinic, 47*(2), 120-124. doi:10.1177/1053451211414188

Clarke, B., Baker, S. K., Smolkowski, K., & Chard, D. (2008). An analysis of early numeracy curriculum-based measurement. *Remedial & Special Education*, 29, 46-57.

Clarke, B., & Shinn, M. R. (2004). A Preliminary Investigation Into the Identification and Development of Early Mathematics Curriculum-Based Measurement. *School Psychology Review*, 33(2), 234-248.

Clements, D. H. (1984). Training effects on the development aneneralization of Piagetian logical operations and knowledge of number. *Journal of Educational Psychology, 76,* 766– 776.

Cobb, P., & Hodge, L. L. (2002). A relational perspective on issues of cultural diversity and equity as they play out in the mathematics classroom. *Mathematical Thinking & Learning, 4*(2), 249-284.

Cockrell, K. S. Placier, P. L., Cockrell, D. H., & Middleton, J. N. (1999). Coming to terms with “diversity” and “multiculturalism” in teacher education: Learning about our students, changing our practice. *Teaching and Teacher Education, 15*(4), 351-366. Common Core State Standards Initiative (2010). Retrieved from: http://www.corestandards.org

Cunningham, James W., and Jill Fitzgerald. "Epistemology and reading." *Reading Research Quarterly* 31.1 (1996): 36-60.

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal Of Learning Disabilities*, 38(4), 324-332.

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science*, *11*(5), 650-654.

Dwyer, C. (1996). Cut scores and testing: Statistics, judgment, truth, and error. *Psychological Assessment*, *8*(4), 360-362. doi:10.1037/1040-3590.8.4.360

Elliott, S. N., Huai, N., & Roach, A. T. (2007). Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology*, *45*(2), 137-161.

Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., Crawford, G., Weaver, W., Howes, C., Ritchie, S., Kraft-Sayre, M., Pianta, R., & Barnett, W., S., (2005). NCEDL's Multi-State Study of Pre-Kindergarten & Study of State-Wide Early Education Programs (SWEEP). NCEDL Working Paper. The University of North Carolina, FPG Child Development Institute, NCEDL.

Elliott, S. N., Huai, N., & Roach, A. T. (2007). Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology, 45*(2), 137-161.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2005) First grade and educational attainment by age 22: A new story. *American Journal of Sociology 110*(5): 1458-1502.

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal Of Emerging Medicine*, *8*(1), 19.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Social Science Research*, *33*(3), 464-497.

Fawcett, T. (2004). ROC Graphs: Notes and Practical Considerations for Researchers. *Pattern Recognition Letters*, *27*(8):882-891.

Fawcett, T., (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874. (Special issue on ROC analysis).

Fischer, B., Köngeter, A., & Hartnegg, K. (2008). Effects of Daily Practice on Subitizing, Visual Counting, and Basic Arithmetic Skills. *Optometry & Vision Development*, *39*(1), 30-34.

Fletcher, J. M., Denton, C., & Francis, D. J. (2005). Validity of Alternative Approaches for the Identification of Learning Disabilities Operationalizing Unexpected Underachievement. *Journal of Learning Disabilities*, *38*(6), 545-552.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice*.

Fuchs, L. S. (2004). Measurement Research. *School psychology review, 33*(2), 188-192.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What why, and how valid is it? *Reading Research Quarterly, 41*, 93–99.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning & Individual Differences, 20*(2), 89-100. doi: 10.1016/j.lindif.2009.09.003

Fuson, K.C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool the number sets test. *Journal of Psychoeducational Assessment*, *27*(3), 265-279.

Gersten, R., & Chard, D. (1999). Number Sense Rethinking Arithmetic Instruction for Students with Mathematical Disabilities. *The Journal of special education*, *33*(1), 18-28.

Gersten, Clarke, Dimino & Rolphus (2010). *ASPENS Research Summary.*

Gersten, R., & Dimino, J. A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41*(1), 99-108.

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

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude, *Journal of Experimental Child Psychology, 76*(2), 104-122.

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

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B. E., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly*, *21*, 431–454.

Fan J, Upadhye S., & Worster A (2006). *Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Experimental Medicine 8*, 19-20.

Fuchs, D., & Fuchs, L.S. (1989). Effects of examiner familiarity on Black, Caucasian, and Hispanic children: A meta-analysis. *Exceptional Children*, *55*, 303-308.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*, 216–228.

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*(2), 79-92.

Fuson, K. C. (1988). Children's counting and concepts of number. New York: Springer-Verlag.

Fuson, K., & Hall, J. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. Ginsburg (Ed.), The development of mathematical thinking (pp. 49-107). New York: Academic Press.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment, 27*, 265-279.

Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Cambridge, MA: Harvard University Press.

Gelman, R., & Meck, E. (1983). Preschooler's counting: Principles before skill. Cognition, *13*,343-359Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of learning disabilities, 38*(4), 293-304.

Ginsburg, H., & Russell, R. (1981). Social class and racial influences on early mathematical thinking. *Monographs of the Society for Research in Child Development*, *46*(6), doi:10.2307/1165946.

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*(4), 237-261.

Griffin, S. (2002). The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies. In J. M. Roher (Ed.), *Mathematical Cognition* (pp. 1 – 32). Greenwich, CT: Information Age Publishing, Inc.

Griffin, S. (1998) Number worlds. Durham, NH: Number Worlds Alliance.

Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3*(1), 1–49.

Griffin, S., Case, R., & Siegler, R. S. (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. 25-49). Cambridge, MA: MITPress/Bradford Books.

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

Hamre, B., & Pianta, R. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School Readiness & the Transition to Kindergarten in the Era of Accountability* (pp. 49-84). Baltimore: Brookes.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*(3), 615-626. doi: 10.1037/0022-0663.93.3.615

Hanley, J. A. (1982). Characteristic (ROC) curve. *Radiology, 743*, 29-36.

Heckert, T. M. (2002). Gender differences in anticipated salary: Role of salary estimates for others, job characteristics, career paths, and job inputs. *Sex Roles, 47*(3), 139.

Henderson, A.R. (1993). Chemistry with confidence: should clinical chemistry require confidence intervals for analytical and other data*?. Clinical Chemistry, 39*, 929-935.

Ilg, F., Ames, L.B. (1951) Developmental trends in arithmetic. *The Journal of Genetic Psychology, 79*, 3.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221-1247.

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

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*(1), 36-46. doi: 10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences, 6*(4), 413 – 432.

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

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1994). Development of calculation abilities in middle- and low income children after formal instruction in school. *Journal of Applied Developmental Psychology*, 15, 223 – 240.

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *Journal of Learning Disabilities, 28,* 53–64.

Jordan, N. C., & Montani, T. O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general. *Journal of Learning Disabilities, 30*(6), 624.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153-175. doi: 10.1111/j.1467-8624.2006.00862.x

Kagan, S. L., & Kauerz, K. (2007). Reaching for the whole: Integration and alignment in early

Klingner, Janette K., and Patricia A. Edwards. "Cultural considerations with response to intervention models." *Reading Research Quarterly* 41.1 (2006): 108-117.

Lembke, E., & Foegen, A. (2009). Identifying Early Numeracy Indicators for Kindergarten and First-Grade Students. *Learning Disabilities Research & Practice*, *24*(1), 12-20. doi:10.1111/j.1540-5826.2008.01273.x

Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. T. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities*, *39*(5), 390-398.

Jitendra, A., & Xin, Y. P. (1997). Mathematical word-problem-solving instruction for students with mild disabilities and students at risk for math failure A research synthesis. *The Journal of Special Education, 30*(4), 412-438.

Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for the development of mathematical understanding: Rational numbers and functions. In S. Carver & D. Klahr (Eds.), *Cognition and instruction* (pp. 1–38). Mahwah, NJ: Erlbaum.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425-461.

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. doi:10.1037/0012-1649.42.1.59

Kintsch, W. & Greeno, J. G. (1985). Understanding and solving arithmetic word problems. *Psychological Review*, 92:109–129.

Ladson-Billings, G. (1997). It doesn't add up: African american students' mathematics. *Journal for Research in Mathematics Education, 28*(6), 697.

Lee, J., Autry, M. M., Fox, J., & Williams, C. (2008). Investigating children's mathematics readiness. *Journal of Research in Childhood Education, 22*(3), 316-328. doi: 10.1080/02568540809594630

Li-Grining, C., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, *46*(5), 1062-1077. doi:10.1037/a0020066.

Linan-Thompson, S., Cirino, P. T., & Vaughn, S. (2007). Determining English language learners' response to intervention: Questions and some answers. *Learning Disability Quarterly*, 185-195.

Locuniak, M.N. & Jordan, N.C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*(5), 451-459.

Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with head start children. *Journal of Educational Psychology, 96*(4), 648-659. doi: 10.1037/0022-0663.96.4.648.

Manly, J. (2004). Advantages and Disadvantages for Separate Norms for African Americans. *The Clinical Neuropsychologist*, *19*, 270- 275.

Mastropieri, M. A., & Scruggs, T. E. (2005). Feasibility and consequences of response to intervention: Examination of the issues and scientific evidence as a model for the identification of individuals with learning disabilities. *Journal of Learning Disabilities, 38*(6), 525–531.

McCardle, P., Mele-McCarthy, J., Cutting, L., Leos, K., & D'Emilio, T. (2005). Learning disabilities in English language learners: Identifying the issues. *Learning Disabilities Research & Practice*, *20*(1), 1-5.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, *40,* 633-645.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*(3), 194.

Moffatt, L., Anderson, A., Anderson, J., & Shapiro, J. (2009). Gender and mathematics at play: Parents' constructions of their pre-schoolers' mathematical capabilities. *Investigations in Mathematics Learning*, *2*, 1-25.

Molfese, V. J., Molfese, D. L., Modglin, A. T., Walker, J., & Neamon, J. (2004). Screening early reading skills in preschool children: Get Ready to Read.*Journal of Psychoeducational Assessment*, *22*(2), 136-150.

Moschkovich, J. N. (1999). Understanding the needs of latino students in reform-oriented mathematics classrooms. *Changing the Faces of Mathematics: Perspectives on Latinos, 4*, 5-12.

Nasir, N.S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics in school: Boundaries between "cultural" and "domain" knowledge in the mathematics classroom and beyond. *Review of Research in Education*, *32*(1), 187-240.

National Center for Education Statistics (2009). *The Nation's Report Card: Mathematics 2009* (NCES 2010–451). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C http://nces.ed.gov/nationsreportcard/pubs/main2009/2010451.asp

Obuchowki, N. (2000). Sample Size Tables For Receiver Operating Characteristic Studies. *American Journal of Roentgenology, 173*(3), 603-608.

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

Ojose, B. (2008). Applying Piaget's Theory of Cognitive Development to Mathematics Instruction. *The Mathematics Educator, 18*(1), 26-30.

Pappas, S., Ginsburg, H., Jiang, M. (2003). SES differences in young children's metacognition in the context of mathematical problem solving. *Cognitive Development*, *18*(3), 431-450.

Paulos, J.A. (1988). Innumeracy: mathematical illiteracy and its consequences. New York: Hill and Wang.

Payne, J. E.(1990). Mathematics for the young child. Reston, VA: National Council of Teachers of Mathematics.

Pennington, B. F., Gilger, J. W., Olson, R. K., & DeFries, J. C. (1992). The External Validity of Age-Versus IQ-Discrepancy Definitions of Reading Disability Lessons From a Twin Study. *Journal of Learning Disabilities*, *25*(9), 562-573.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, 79, 375 – 394.

Ray, K., & Smith, M. C. (2010). The kindergarten child: What teachers and administrators need to know to promote academic success in all children. *Early Childhood Education Journal, 38*(1), 5-18. doi: 10.1007/s10643-010-0383-3

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*(5), 615.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P.Ginsburg (Ed.) , The development of mathematical thinking (pp. 153– 196). New York: Academic Press.

Rinaldi, C., & Samson, J. (2008). English Language Learners and Response to Intervention Referral Considerations. *Teaching exceptional children*, *40*(5), 6-14.

Ritchie, S., Maxwell, K., & Clifford, R. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School Readiness & the Transition to Kindergarten in the Era of Accountability* (pp. 85-96). Baltimore: Brookes.

Rodriguez, (2005). Teachers' Resistance to Ideological and Pedagogical Change: Definitions, Theoretical Framework, and Significance. In A.J. Rodriguez and R. Kitchen (Eds.), *Preparing Prospective Mathematics and Science Teachers to Teach for Diversity: Promising Strategies for Transformative Pedagogy*, (pp. 1-16) Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rotello, C., Masson, M., & Verde, M. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics,* 70*(2),* 389-401.

Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition & Instruction, 1*(2), 217.

Russell, J., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., Ross, C., Schochet, P. and Swank, P. (2011;2007). National Evaluation of Early Reading First: Final Report. (NCEE 2007-4007rev). U.S. Department of Education, Institute of Education Sciences, Washington, DC: U.S. Government Printing Office.

Sarnecka, B., & Carey, S. (2008). How Counting Represents Number: What Children Must Learn and When They Learn It. *Cognition, 108*(3), 662-674.

Saxe, G. B., & Kaplan, R. G. (1981). Gesture in early counting: A developmental analysis. *Perceptual and Motor Skills,* 53 (85 l-854.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18*(1), 253-286. Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*(1), 37-59.

Shaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, *6*, 357-379.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444. doi: 10.1111/j.1467-8624.2004.00684.x

Simmons, D. C., Coyne, M. D., Kwok, O., McDonagh, S., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention a longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*(2), 158-173.

Smith, L. (2002). *Reasoning by mathematical induction in children's arithmetic.* (Oxford, Pergamon Press).

Starkey, P. & Klein, A. (2000). Fostering Parental Support for Children's Mathematical Development: An Intervention with Head Start Families. *Early Education & Development*, *11*(5), 659-680. doi:10.1207/s15566935eed1105_7

Squire, K. D., MaKinster, J. G., Barnett, M., Luehmann, A. L., & Barab, S. L. (2003). Designed curriculum and local culture: Acknowledging the primacy of classroom culture. *Science education*, *87*(4), 468-489.

Stock, P., Desoete, A., & Roeyers, H. (2009). Mastery of the counting principles in toddlers: A crucial step in the development of budding arithmetic abilities*? Learning & Individual Differences*, *19*(4), 419-422.

Stock, P., Desoete, A., & Roeyers, H. (2010). Detecting Children with Arithmetic Disabilities From Kindergarten: Evidence From a 3-Year Longitudinal on the Role of Preparatory Arithmetic Abilities. *Journal of Learning Disabilities*, *43*, 250-268.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, *80,* 217–222.

Starkey, P. & Klein, A. (2000). Fostering Parental Support for Children's Mathematical Development: An Intervention with Head Start Families. *Early Education & Development*, *11*(5), 659-680. doi:10.1207/s15566935eed1105_7

Swanson, L., & Kim, K. (2007). Working Memory, Short-Term Memory, and Naming Speed as Predictors of Children's Mathematical Performance. *Intelligence*, *35*(2), 151-168.

Swets, J. (1988). Measuring the Accuracy of Diagnostic Systems. *Science*, *240*, 1285–1293.

Teisl, J. T., Mazzocco, M. M., & Myers, G. F. (2001). The Utility of Kindergarten Teacher Ratings for Predicting Low Academic Achievement in First Grade. Journal of Learning Disabilities, *34*(3), 286.

VanDerHeyden, A.M. (n.d.) Using RTI to Improve Learning in Mathematics. Retrieved from: http://www.rtinetwork.org/learn/what/rtiandmath

VanDerHeyden, A. M., Broussard, C., Fabre, M., Stanley, J., & al, e. (2004). Development and validation of curriculum-based measures of math performance for preschool children. *Journal of Early Intervention*, *27*(1), 27-29, 31-41.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning disabilities research & practice*, *18*(3), 137-146.

van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology, 101*(2), 99-113.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*(2), 157-169.

Vygotsky, L.S. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Wells, A. S., & Serna, I. (1996). The politics of culture: Understanding local political resistance school for children from low-income families. *Journal of Educational Psychology*, to detracking in racially mixed schools. *Haruard Educational Review,66,* 93-118.

Welsh, J., Nix, R., Blair, C., Bierman, K., & Nelson, K. (2010). The development of cognitive skills and gains in academic *102*(1), 43-53. doi:10.1037/a0016738.

Willis, G.B. & Fuson, K.C. (1988). Teaching children to use semantic drawings to solve addition and subtraction word problems. *Journal of Educational Psychology, 80*, 192-201.

Witzel, B. S., & Riccomini, P. J. (2007). Optimizing math curriculum to meet the learning needs of students. *Preventing School Failure, 52*(1), 13-18.

Worster, A., Fan, J., & Upadhye, S. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Experimental Medicine*, *8*(1), 19-20.

Wynn, K. (August 1992). “Addition and Subtraction by Human Infants.” *Nature, 358*(6389), 749–750.

Xu, Y., & Drame, E. (2008). Culturally appropriate context: Unlocking the potential of response to intervention for English language learners. *Early Childhood Education Journal*, *35*(4), 305-311.

Ysseldyke, J., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an Instructional Management System to Improve Mathematics Skills for Students in Title I Programs. *Preventing School Failure*, *48*(4), 10-14.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology, 110*(4), 481-498.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, *39*(4), 561-577.

Appendix A

PARENTAL PERMISSION FORM
Math Study

I agree to allow my child, ____________________, to take part in a research study titled, "Practicing Prevention: Early Mathematics Screening in Kindergarten" which is being conducted by Dr. Martha Carr and Sara Woodruff, from the Educational Psychology and Instructional Technology Department at the University of Georgia (xxx-xxx-xxxx). I do not have to allow my child to be in this study if I do not want to. My child can refuse to take part at any time without giving any reason, and without penalty or loss of benefits to which my child is otherwise entitled. I can ask to have the information related to my child returned to me, removed from the research records, or destroyed. Participation is voluntary.

- **The reason for the study is to find better ways to teach children mathematics and to improve assessments for children in kindergarten.**

- **Children who take part in this research may provide their teachers with additional information to improve instruction based on student strengths and weaknesses (as revealed in the assessments) students will complete.**

- **If I allow my child to take part, my child will work with a researcher at three times: January or February and April or May of kindergarten. Demographic information will be obtained from school records (including race/ethnicity). Students will be reassessed in August or September of first grade. This will require approximately 30 minutes each time. During each session, students will complete brief math assessments.**

- **The research is not expected to cause any harm. Minimal discomfort may be experienced in the form of fatigue. Breaks will be given as needed to address fatigue as needed.**

- **Any individually-identifiable information collected about my child will be held confidential unless otherwise required by law. My child's identity will be coded, which links the data and all data will be kept in a secured location. The code will be destroyed upon the completion of the study.**

- **I understand the study procedures described above. My questions have been answered to my satisfaction, and I agree to allow my child to take part in this study. Additional questions will be addressed at any time during the study. I have been given a copy of this form to keep.**

Sara Woodruff Email: saraew10@uga.edu; Phone: (xxx) xxx-xxxx
Martha Carr Email: mmcarr@uga.edu

_________________________ _________________________ ____
Name of Parent or Guardian Signature Date

Child' date of birth________________ Child's teacher______________________

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your child's rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 629 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu

Appendix B

Number line/number magnitude. There are several approaches to understanding the evolution of children's representation of number. Generally, children tend to represent number either linearly or logarithmically (Siegler & Booth, 2004). When children use logarithmic representations, small numbers (i.e. 8 and 9) are more widely spaced apart in the number line and large numbers are compressed (Fisher & Campenas, 2009). Children may find the logarithmic representation useful when estimating unfamiliar quantities, particularly because it exaggerates the difference between numbers in the higher ranges and allows the child to discriminate more accurately between numbers in the higher range (Siegler & Booth, 2004). Consistent with this finding, discrimination between quantities that are similar in magnitude will result in more overlap and prove more difficult to discriminate than quantities of greater magnitude (Siegler & Opfer, 2003).