Georgia State University
ScholarWorks @ Georgia State University

Psychology Dissertations | Department of Psychology

Spring 5-10-2014

Structure of Mathematics Acheivement and Response to Intervention in Children with Mild Disabilities

Matthew Foster
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/psych_diss

Recommended Citation
Foster, Matthew, "Structure of Mathematics Acheivement and Response to Intervention in Children with Mild Disabilities." Dissertation, Georgia State University, 2014.
https://scholarworks.gsu.edu/psych_diss/127

This Dissertation is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

STRUCTURE OF MATHEMATICS ACHIEVEMENT AND RESPONSE TO INTERVENTION IN CHILDREN WITH MILD DISABILITIES

by

MATTHEW E. FOSTER

Under the Direction of Dr. Rose A. Sevcik

ABSTRACT

Children with mild disabilities are known to have difficulties with developing mathematical skills (Hoard, Geary, & Hamson, 1999). Yet, children with mild *intellectual* disabilities (MIDs) have rarely been included in rigorous scientific research. The present study has three goals. The first goal was to determine the structure of mathematics achievement in elementary aged children with MIDs and children with reading disabilities (RDs) without accompanying mathematics disabilities. The second goal was to establish the measurement stability of mathematics achievement. The third goal was to evaluate students' response to a mathematics intervention. The participants were 265 children with MIDs and 137 children with RDs. Confirmatory factor analysis and measurement invariance evaluation was utilized to determine the structure of mathematics achievement and to ensure reliable and valid measurement of mathematics achievement between groups across three time points. The results of measurement invariance evaluation indicated that a joint model specification, characterized by two groups, both of which included children with MIDs and children with RDs who were differentiated according to intervention condition participation (not disability status), provided the best account of the underlying data structure. Further, the structure of mathematics achievement in the present sample was unidimensional, and the measurement of mathematics achievement was temporally stable between groups. Finally, latent mathematics achievement growth was evaluated. The results indicated that students in the mathematics intervention condition evidenced an advantage over those in a reading intervention condition at mid- and post-intervention evaluation, while also evidencing more growth in this conceptual domain. Instructional implications are discussed in terms of topic choice and pacing.

INDEX WORDS: Mathematics, Math achievement, Mild disabilities, Mild intellectual disabilities, Learning disabilities, Reading disabilities, Measurement invariance, Measurement equivalence, Mathematics achievement structure, Mathematics achievement nature, Response to intervention, Math intervention.

STRUCTURE OF MATHEMATICS ACHIEVEMENT AND RESPONSE TO INTERVENTION IN CHILDREN WITH MILD DISABILITIES

By

MATTHEW E. FOSTER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2014

Copyright by
Matthew E. Foster
2014

STRUCTURE OF MATHEMATICS ACHIEVEMENT AND RESPONSE TO INTERVENTION IN CHILDREN WITH MILD DISABILITIES

By

MATTHEW E. FOSTER

Committee Chair: Rose A. Sevcik

Committee: Robin D. Morris

Lee Branum-Martin

Paul T. Cirino

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2014

DEDICATION

To my loving parents, Walter and Deborah Foster, thank you for your love, encouragement, and support. I am extremely fortunate to have you as parents. Without you, I would not have accomplished this goal.

ACKNOWLEDGEMENTS

I wish to express my appreciation to Dr. Rose A. Sevcik, my faculty supervisor, under whose direction this dissertation was conducted. Her support and guidance was a source of strength throughout the dissertation process. Thank you Rose, for the guidance that you have provided me over the last several years.

I am indebted to Dr. Lee Branum-Martin for his guidance and valuable feedback throughout the completion of this dissertation. Thank you Lee, for sharing your understanding and excitement for measurement; you have made a lasting influence on my methodological approach to research.

Particular appreciation is extended to Dr. Robin Morris, who has provided me ongoing support during the completion of this dissertation as well as throughout my doctoral training. Robin, the mentorship that you have provided me at each phase of my doctoral training is greatly appreciated.

To Dr. Paul Cirino, I wish to express my thanks and gratitude for his support and valuable feedback throughout the dissertation process. Paul, thank you for collaborating on this study and sharing your expertise.

Additionally, I would like to thank Dr. Maryann Romski, who has provided me with mentorship throughout my doctoral training. Maryann, your support and guidance has helped me achieve several milestones during my training and contributed to my completion of the Developmental Psychology doctoral training program.

Finally, I would like to thank the Institute for Education Sciences and the National Institutes for Health, who awarded grants to Drs. Sevcik (#H324K040007) and Morris (#R01 HD30907). Without their support and that of the school districts, administrators, educators, and participants, this dissertation would not be possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

1. INTRODUCTION 1

1.1 The Importance of Mathematics Achievement 1

1.2 The Nature of Mathematics Achievement in Elementary School Children 3

1.3 Children with Mild Intellectual Disability and Children with Specific Learning Disability 4

1.4 Mild Disabilities and Mathematics Achievement: Empirical Study 6

1.5 Response to Mathematics Interventions 7

1.6 Overview of the Present Study 12

1.7 Research Aims 16

1.8 Research Questions 18

2. METHOD 20

2.1 Participants 20

2.2 Assessment Instruments 22

2.3 Mathematics and Reading Intervention Programs 23

2.4 Procedure 25

3. RESULTS 26

3.1 Descriptive Statistics 26

3.2 Missing Data Patterns 30

3.3 Confirmatory Factor Analysis (CFA) for Mathematics Achievement 31

3.4 Data Preparation and Special Considerations 31

3.5 Testing the Structure of Mathematics Achievement 32

3.6 Evaluating Students' Response to Mathematics Intervention 50

3.7 Treatment Effects 52

3.8 Differential Indicator Bias 54

4. DISCUSSION....58

4.1 The Structure of Mathematics Achievement.... 58

4.2 Response to Mathematics Intervention....61

4.3 Language Concerns in the Measurement of Mathematics Achievement....65

4.4 Instructional Implications....66

4.5 Limitations and Future Directions....67

4.6 Conclusion....69

5. REFERENCES....71

6. APPENDICES....81

Appendix 1. Descriptive Statistics by Disability Group....81

Appendix 2. Descriptive Statistics for Combined Disability Groups....84

Appendix 3. Within Group Longitudinal Measurement Invariance Evaluation....87

Appendix 4. Unstandardized Latent and Residual Correlations with Standard Deviations....89

LIST OF TABLES

Table 1. Descriptive Statistics by Intervention Group across Studies: Pre-intervention..............................28

Table 2. Descriptive Statistics by Intervention Group across studies: Mid-intervention.............................29

Table 3. Descriptive Statistics by Intervention Group across Studies: Post-intervention............................30

Table 4. Measurement Invariance Evaluation for Mathematics Achievement: Combined Group Model..35

Table 5. Measurement Invariance Evaluation for Mathematics Achievement: Joint Group Model............40

Table 6. Joint Group Model: Final Model Parameter Estimates..46

Table 7. Pre- and Post-Intervention Means, Standard Deviations, and Effect Sizes Across Studies...........54

Table 8. Predicted Observed Scores by Group for Non-Invariant Indicators at Post-Intervention.............56

LIST OF FIGURES

Figure 1. Combined Group Model: Equal Form..34

Figure 2. Combined Group Model: Equivalent Factor Loadings..36

Figure 3. Combined Group Model: Equivalent Intercepts..37

Figure 4. Joint Group Model: Equal Form..39

Figure 5. Joint Group Model: Equivalent Factor Loadings...42

Figure 6. Joint Group Model: Equivalent Intercepts...44

Figure 7. Addition Intercepts by Intervention Group..47

Figure 8. Measurement Intercepts by Intervention Group...47

Figure 9. Time/Money Intercepts by Intervention Group..48

Figure 10. Longitudinal Mathematics Achievement by Intervention Group..52

Figure 11. Predicted Post-Intervention Observed Scores for Addition by Intervention Group...................56

Figure 12. Predicted Post-Intervention Observed Scores for Time/Money by Intervention Group............57

1. INTRODUCTION

1.1 The Importance of Mathematics Achievement

In high wage industrialized countries such as the United States, mathematics underachievement is related to poor educational and occupational success. For instance, middle school math proficiency is related to enrollment in advanced high school math courses, which is subsequently, related to an increased likelihood of graduating from college (National Math Advisory Panel, 2008; Sadler & Tai, 2007) and increased employability, productivity, and wages (Altonji, 1995; Joensen & Nielsen, 2009; Riveria-Batiz, 1992) 10 years after high school graduation (Rose & Betts, 2004). Further, the influence of mathematics achievement on occupational success is robust. Using the High School and Beyond data set, Rose and Betts (2004) demonstrated that the influence of high school mathematics achievement on occupational success remained significant even after accounting for the influence of a multitude of covariates. These covariates included the individual's demographic (e.g., race/ethnicity, gender, age, marital status), family (e.g., parental income, education, number of siblings) and school (e.g., student-teacher ratio, books per pupil, length of school year, school enrollment, average spending per student, geographic region) characteristics as well as the individual's highest educational degree attained, college major, and occupation. In short, "math matters" (Rose & Betts, 2004; p. 501); however, children with mild disabilities (i.e., mild intellectual disabilities—MIDs and learning disabilities—LDs) may be less likely to experience increased employability and earnings due in part to a lack of enrollment in advanced high school mathematics courses. Unfortunately, Rose and Betts (2004) could not directly examine the influence of high school mathematics achievement on occupational success for students with MIDs as demographic information for MID was not included in the High School and Beyond data set.

Early research suggests that mathematics achievement in children with MIDs and children with LDs (i.e., RD, mathematics disability-MD, and RDMD) lags behind that of their typically achieving peers. For instance, Cawley and Miller (1989) and Cawley et al. (2001) identified that students with mild disabilities (i.e., MID and LD) required between two and three years of schooling to show one year of academic progress (Cawley & Miller, 1989; Cawley et al., 2001). Consequently, on average, mathematics achievement of students with mild disabilities, upon exiting school, is near the fifth or sixth grade competency level (Warner et al., 1980). In contrast to this early research, more recent investigations documenting the academic difficulties of children with mild disabilities have focused on high school dropout rates (e.g., Polloway, Lubin, Smith, & Patton, 2010) and post school outcomes (Cameto, 2005). According to the President's Commission on Excellence in Special Education (2002), children identified as having a disability (i.e., any of the 13 disability categories) are twice as likely to drop out of school compared to their non-disabled peers; with 29% of students with intellectual disabilities and 32% of students with LDs dropping out of school (Polloway et al., 2010). In regard to post school outcomes, Cameto (2005) utilized the National Longitudinal Transition Study-2 (NLTS-2) data and identified that 25% of individuals with intellectual disabilities and 46% of individuals with LDs were employed one-to-two years following their graduation compared to 55% (42% who were going to college were employed and 78% of those were not were employed) of recent high school graduates from the general population. However, Cameto (2005) did not fully examine the influence of enrollment in post secondary school experiences on employment rates for individuals with disabilities. Therefore, the employment rates of individuals with intellectual disabilities and those with LDs should be interpreted cautiously. Despite this caution, the collective results suggest that students with mild disabilities are at risk for poor educational achievement and consequently, meager employment outcomes.

In order to foster occupational success, it is important to support academic achievement and in particular, mathematics achievement, early in the lives of students with mild disabilities. Improving mathematics achievement may improve high school graduation rates, college enrollment and graduation rates, and substantially improve occupational success (e.g., reduce unemployment and underemployment, increase full time employment, increase wages) in individuals with MIDs and LDs. To do so, empirical study related to improving mathematics achievement in children with mild disabilities should be furthered. In particular, this research field has discussed the possibility of quantitative verses qualitative differences between children with MIDs and children with LDs for a few decades (e.g., Parmar, Cawley, & Miller, 1994); however, the structure (or nature) of mathematics achievement as it relates to these special populations has yet to be systematically investigated using rigorous statistical methodology. Further, studies concerned with students' response to mathematics interventions have been limited. In particular, empirical work that has included children with MIDs and children with LDs within the same empirical study are often characterized by small samples and fail to establish between group longitudinal measurement invariance, which is a precondition to studying group differences and longitudinal change. The present study will therefore systematically investigate the nature of mathematics achievement in these special populations and establish equivalent measurement of mathematics achievement before investigating students' response to a mathematics intervention.

1.2 The Nature of Mathematics Achievement in Elementary School Children

A substantial portion of the research concerned with the development of mathematical competencies in elementary school aged children has focused on arithmetic calculations (Fuchs, Fuchs, & Prentice, 2004; Gersten, Jordan, & Flojo, 2005); however, during these academic years, mathematics is broader than this single area of study. For instance, the Common Core State Standards for Mathematics (CCSSM; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) is an attempt to improve mathematics achievement in the United States by providing a more focused and coherent set of mathematics standards by grade level. The set of standards proposed by the CCSSM (2010) outline in detail, competencies across several areas of mathematics (e.g., numeration, estimation and measurement, problem solving, geometry, and conceptual knowledge) that elementary school students are expected to demonstrate proficiency. Broadly, proficiency related to *numeration* extends beyond small quantities that characterized earlier grades (i.e., kindergarten through first grade) to working with groups of objects to gain foundations for multiplication. Competencies related to *estimation* and *measurement* involve standard units; solving problems involving intervals of time, liquid volumes, and masses of objects; converting measurements from a larger unit to a smaller unit; and converting like measurement units within a given measurement system. *Problem solving* related competencies include representing and solving problems involving addition, subtraction, multiplication, and division; performing the four operations with multi-digit whole numbers and with decimals to the hundredths place value position; and representing and interpreting data. Skills related to *geometry* include reasoning with shapes and their attributes (e.g., number of angles, sides, perimeter, radius, area); classifying shapes by properties of their lines and angles; and graphing points on the coordinate plane to solve problems. Finally, *conceptual knowledge* is an area that involves understanding place value, relationships among the four arithmetic operations, fractions, and geometric concepts. In short, the CCSSM (2010) standards highlight the breadth of areas and skills that elementary school mathematics achievement encompasses.

1.3 Children with Mild Intellectual Disability and Children with Specific Learning Disability

Children with mild intellectual disability (MID) and children with specific learning disability (SLD) are two special populations that are included in a category referred to as ‘mild disabilities.’ However, the umbrella term, *mild*, should not be taken lightly. Impairments associated with each disability are life-long and can affect all areas of an individual’s life (e.g., academic, occupation, social-emotional).

Mild Intellectual Disability (MID)

Individuals with MID evidence significant limitations in both intellectual functioning (reasoning, learning, and problem solving) and adaptive behavior (conceptual, social, and practical adaptive skills), that originate before 18 years of age (American Association on Intellectual Disabilities, 2010). Significant limitations are defined as IQ and adaptive behavior scores that are at least two standard deviations below the population mean (e.g., ≤ 70), with the IQ range for MID being between 55 and 70. In regard to the prevalence of intellectual disability, data from the Metropolitan Atlanta Developmental Disabilities Surveillance Program suggests that 11.7 per 1,000, 8-year-old children have an intellectual disability (Obi et al., 2011). Of those with intellectual disability, early work (Glass, Christiansen, & Christiansen, 1982) suggested that MID accounted for between 75 and 80% of all children diagnosed with intellectual disabilities. With respect to etiology of intellectual disabilities, more recent work suggests that the cause is unknown in 52 (Heikura et al., 2005) to 80% (Rauch et al., 2006) of individuals. When the cause of intellectual disability is known, the leading etiological factors have included Down syndrome, Williams syndrome, Fragile-X syndrome, Cohen syndrome, and monosomy 1p36 (Heikura et al., 2005; Rauch et al., 2006).

Specific Learning Disability (SLD)

A specific learning disability (SLD) is a disorder in one or more basic cognitive processes (e.g., input, integration, memory, output, and motor) involved in understanding or using spoken or written language that may manifest itself in difficulty with listening, speaking, reading, writing, spelling, or completing mathematical calculations (20 U.S.C. Section 1401(30)). With respect to specific reading (RDs) and mathematics disabilities (MDs), previous versions of the Diagnostic Statistical Manual (DSM) (e.g., DSM-IV TR; American Psychiatric Association, 2000) referred to each as a disorder (i.e., reading disorder, mathematics disorder) where achievement in the respective area is substantially below the individual's expected level, given his or her chronological age, IQ and age appropriate education. In contrast, DSM-V (American Psychiatric Association, 2013) refers to Specific Learning Disorder as a single category of disability (or overall diagnosis) that incorporates deficits that impact academic achievement while providing specifiers for the areas of reading, mathematics, and written expression.

In regard to prevalence estimates, Landerl and Moll (2010) recently utilized a strict (-1.5 SD below age norm) and a lenient (-1. SD below age norm) criterion to identify elementary aged children who exhibited a reading or arithmetic disorder. Their results suggested that between 7.0 (strict criterion) and 14.8% (lenient criterion) of elementary aged children evidence a reading disorder; whereas between 6.1 (strict criterion) and 15.4% (lenient criterion) evidenced a arithmetic disorder. Further, comorbidities between reading and arithmetic disorders also were determined. Of the children with reading deficits that met the lenient criteria, 38.8% presented a comorbid arithmetic deficit, whereas, of those who met the strict criteria, 22.7% also evidenced an arithmetic deficit. Of the children with arithmetic deficits that met the lenient criteria, 37.3% presented a comorbid reading deficit, whereas, of those who met the strict criteria, 25.9% also evidenced a comorbid reading deficit. In short, it appears that it is relatively common for children with a reading or arithmetic deficit to evidence a deficit in the other academic domain.

In regard to the etiology of SLDs, the National Joint Committee on Learning Disabilities (1991) has maintained the position that the basis of learning disorders is presumed to be due to central nervous system dysfunction. Nervous system dysfunction in children with SLDs may be characterized by different activation patterns during phonological processing tasks, for example, compared to nondisabled children (Miller, Sanchez, & Hind, 2003; Simos, et al., 2000). Other implicated causes of SLDs include genetics/heredity, tobacco, alcohol, and other drug use during pregnancy, complications during pregnancy, environmental toxins, poor nutrition, and maturational delay (Pierangelo & Giuliani, 2007).

1.4 Mild Disabilities and Mathematics Achievement: Empirical Study

Children with mild disabilities are known to have difficulties developing mathematical skills (Hoard, Geary, & Hamson, 1999). Subsequently, two groups of children with mild disabilities, those with MID and their peers with SLD, exit school with poor mathematics proficiency (Warner et al., 1980). However, empirical study of mathematics achievement has largely focused on employing models of typical development in understanding the mathematics development of children with SLDs, and in particular, children with MD. As a consequence, sufficiently well-developed theoretical models and experimental techniques have been developed to guide the study of mathematics development and achievement in children with SLDs (Geary, Hoard, & Hamson, 1999). In contrast, few studies have included children with MIDs, and consequently, very little is known about the nature of mathematics achievement in children from this special population. It is therefore valuable to include children with MIDs in studies concerned with mathematics skill development and achievement.

Early research suggested that there is a disparity between the mathematics skill sets characteristic of students with MIDs compared to that of their peers with LDs. For instance, Parmar et al. (1994) investigated differences in mathematics performance and rate of skill growth in 206 students with mild mental retardation (MMR) and 295 students with LD (students' specific area of disability [i.e., MD, RD, MDRD] was not described) between the ages of 8 and 14 years. Skill performance was individually assessed across four mathematical domains (i.e., Basic Concepts, Listening Vocabulary, Problem Solving, Fractions). The results suggested that students with LD evidenced significantly higher mean scores across each domain. Parmar et al. (1994) therefore concluded that students with LDs demonstrated greater growth rates than their age-equivalent peers with MMR, noting that students with MMR at the highest age group (14 years) were unable to achieve, on average, as younger students (8-, 9-, and 10-year-olds) with LDs. As a consequence, Parmar et al. (1994) inferred that the nature of mathematics achievement was different in children with MMR and children with LDs.

1.5 Response to Mathematics Interventions

Mathematics difficulties may be related to a particular skill area (e.g., numeration, estimation and measurement, problem-solving, geometry, and conceptual knowledge.) or they can be more severe, affecting several different areas (Kroesbergen & Van Luit, 2003). The potential causes for these difficulties are numerous; however, poor fit between the learning characteristics of individual students and the instruction they receive is a likely cause (Carnine, 1997). Subsequently, different intervention methods (e.g., direct instruction, strategy instruction, computer assisted instruction) have been employed to remediate ‘prerequisite skills’ (i.e., counting and number sense), ‘basic skills’ (i.e., arithmetic facts), and the use of mathematical ‘problem-solving strategies’ in low performing/at risk students and students with mild disabilities (Kroesbergen & Van Luit, 2003). Intervention study outcomes carry the potential to inform teaching pedagogy and ameliorate mathematics difficulties in children with and without disabilities.

Through the provision of theoretically informed, evidence-based instructional practices, children with disabilities may be more likely to gain essential mathematical knowledge and skills during their elementary school years. As a result, children with disabilities may be more highly motivated to enroll in more advanced high school mathematics courses that are related to several positive outcomes as a young adult (see National Mathematics Advisory Panel, 2008; Rose & Betts, 2004; and Sadler & Tai, 2007). To date, research concerned with improving mathematics skills has rarely included children with MIDs and children with SLDs within the same study; however, some limited empirical work exists that has included both special populations. Such studies have included interventions that targeted early numeracy skills (e.g., Van Luit & Schopman, 2000) arithmetic calculation skills, and arithmetic facts and coin sums (e.g., Mattingly & Bott, 1990; Miller & Mercer, 1993; Podell, Tournaki-Rein, & Lin, 1992; Van Luit, 1987; 1994; Van Luit & Naglieri, 1992).

Early Numeracy Intervention

The acquisition of early numeracy skills (e.g., subitizing, rote counting, enumeration, counting procedures, and concepts of comparison, classification, seriation, and correspondence) is crucial to the development of basic arithmetic skills involving the four operations (i.e., addition, subtraction, multiplication, division). For instance, Van Luit and Schopman (2000) identified 124 Dutch students between the ages of five and seven years as, ‘low mathematics achievers’ (i.e., children with a score comparable to the lowest 25% of the norm group on an early numeracy norm-referenced test), which included students with MIDs and students with LDs. Students were assigned to an experimental (n = 62) or comparison (n = 62) group, matched for gender, age, and early numeracy performance. Mean (and standard deviation) age in years and IQ of the children in the experimental group was 6.30 (0.5) and 74.90 (13.1), respectively. For the children in the comparison group, mean (*sd*) age in years and IQ (*sd*) was 6.10 (0.5) and 79.10 (14.3), respectively. The early numeracy intervention consisted of twenty 30-minute instructional lessons (focusing on numbers between 1 and 15) that were delivered to small groups of students twice a week, and alternated between the use of concrete, semiconcrete, and abstract representations of number. Results showed that the intervention group significantly improved with respect to several early numeracy skills (e.g., comparison, using number names, general understanding of number, and several types of counting procedures). Moreover, in comparing the effect size (Cohen’s *d*) for the intervention group (1.44) with that of the comparison group (0.68), the result suggests that children in the former group scored higher than those in the latter group on the outcome measures at posttest assessment.

Arithmetic Calculation Skills, Arithmetic Facts and Coin Sums Interventions

Knowledge of arithmetic facts appears to be a part of the foundation for later mathematics learning (National Mathematics Advisory Panel, 2008) while also being a source of mathematics difficulties (e.g., Geary et al., 1999). Early research has provided mixed findings in regard to the effectiveness of various interventions in improving students’ basic mathematics skills. In one attempt, Mattingly and Bott (1990) utilized a constant time delay procedure with a multiple-probe design to facilitate multiplication fact acquisition in two fifth and two sixth grade students (age range: 11-12 years). Of the four students, two were identified as educable mentally handicapped (IQ = 65 and 71, respectively), one as learning disabled (IQ = 101), and one as evidencing a behavior disorder (IQ = 91). The results suggested that the four students learned a set of 30 multiplication facts; although the number of minutes and number of one-on-one direct instruction sessions required for learning the facts to criterion (i.e., 100% accuracy for three consecutive instructional sessions), varied across the students, requiring between 280 and 388 minutes and between 86 and 111 sessions. Further, mean rate of correct responding following implementation of the intervention for all responses, across all students, was 98.3%. It therefore appears that the introduction of the time delay procedure (or intervention) accelerated correct responding for multiplication facts that the students had not yet learned.

In another study (Miller & Mercer, 1993) concerned with improving students' proficiency with solving arithmetic facts, a concrete-semiconcrete-abstract teaching sequence (Miller, Mercer, & Dillon, 1992) was utilized within the context of a multiple baseline across subjects design, for three separate investigations. Participants included nine elementary school students; five were classified as students with SLDs (age range: 7.7-9.7 years; IQ: 71-85), three as at risk for a SLD (age range: 10.1-11.3 years; IQ not reported), and one as educably mentally handicapped (EMH) who was 8.3 years of age with an IQ score of 63. The three areas of arithmetic facts (addition facts; division facts; coin sums) were taught to different students such that *addition facts* were taught to the three students with SLD; *division facts* were taught to the three students at risk for SLD; and *coins sums* were taught to the two students with SLD and the student identified as EMH. The instructional sessions were 20 minutes in duration and consisted of providing students with an advanced organizer, demonstration of the skill followed by student modeling of the skill, provision of guided practice with feedback, and finally, independent practice. The results of Miller and Mercer (1993) suggested that between three and seven lessons (i.e., 60-140 minutes of instruction) using manipulative devices (i.e., concrete phase) and pictures (i.e., semiconcrete phase) were needed, before students transferred their learning to abstract type problems in each of the domains, respectively. Thus, transfer to accurately solving abstract problems within each of the three areas (i.e., addition and division facts and coin sums) required between 6 and 14 lessons (i.e., 120-280 instructional minutes). From the report, it was not possible to determine whether the teacher provided one-on-one instruction or instruction to the groups of three students at once.

In a study concerned with the effects of computer assisted instruction (CAI), Podell et al. (1992) compared the effects of CAI with that of a traditional approach (i.e., utilized worksheets and provided positive reinforcement and corrective feedback) in promoting automatization of basic addition and subtraction skills. The addition and subtraction interventions were separate investigations. The former investigation included 52 students, whereas the latter included 42. With respect to the addition investigation, 24 students were described as non-handicapped second graders, M_{age} (*sd*) = 7.78 (0.74) years. The other 28 participants were described as second to fourth grade students with mild mental handicaps, M_{age} (*sd*) = 8.62 (1.51) years, and included children with MMR and LDs. For the subtraction investigation, 20 students were described as non-handicapped, M_{age} (*sd*) = 7.79 (0.80) years, while 22 were second to fourth grade students that evidenced mild mental handicaps, M_{age} (*sd*) = 8.82 (1.53) years, and again, included children with MMR and LDs. Students with and without disabilities were randomly assigned to the CAI addition (non-handicapped, *n* = 15; mild mental handicaps, *n* = 18) and subtraction (non-handicapped, *n* = 14; mild mental handicaps, *n* = 14) intervention conditions. CAI was provided using the Math Blaster computer program, which was described as providing instruction via drill-and-practice. Further, the Math Blaster program had an authoring capability, a built-in scoring mechanism and a timer. The authoring feature enabled the researchers to create the addition and subtraction programs that gradually increased in difficulty over the course of the intervention. The intervention was considered complete either when students attained mastery of all lessons, or after they had participated in ten 15-minute instructional sessions, whichever came first. For problem solving accuracy, the results did not suggest that the CAI was more effective than the traditional approach. However, students that participated in the CAI intervention reached the accuracy criterion of at least 90% (i.e., 18 of 20 items) in fewer lessons than those in the comparison group. Unfortunately, this treatment effect was limited to the non-handicapped group suggesting that students with mild mental handicaps needed more practice than their typically achieving peers to achieve fluency with basic arithmetic facts.

Finally, Van Luit and Naglieri (1999) assessed the effectiveness of self-instruction methods to increase the use of strategies when solving multiplication and division problems. Their participants included 42 Dutch students with MMR (M_{age} = 12-years-8-months; M_{IQ} = 70.3) and 42 students with LDs (M_{age} = 10-years-10-months; M_{IQ} = 98.5). Each group of children was randomly assigned to the strategy intervention (MMR, n = 21; LD, n = 21) or the comparison condition (MMR, n = 21; LD, n = 21). Children assigned to the strategy intervention received small group instruction (three to six students) for 45-minutes three times a week during a four-month period. The goal of the strategy intervention was to help children use simple multiplication and division results in more complex problems such as 8 × 13 (8 × 10 + 8 × 3) or 64 ÷ 4 (40 ÷ 4 + 24 ÷ 4). The results suggested that children in both experimental conditions improved their accuracy in solving multiplication and division problems. Children who participated in the strategy intervention, however, improved more than their peers as evidenced by an effect size (Cohen's *d*) of 3.45 verses 0.76 for the comparison condition. Further decomposition of the results according to disability status (i.e., MMR and LD) and experimental condition (i.e., intervention and comparison condition) indicated that children with MMR and children with LDs who participated in the strategy intervention demonstrated greater improvement than their disability similar peers as evidenced by higher post-test scores despite nonsignificant pre-test score differences (i.e., pre-intervention test scores were similar). Moreover, of the four groups (i.e., disability status by experimental condition) children with MMR that participated in the strategy instruction intervention outperformed children with LDs that participated in the comparison condition at the post-test evaluation. This finding suggests that the strategy intervention was effective in helping children with MMR catch up to their peers (albeit slightly younger peers) who evidence less severe cognitive impairments as characterized by IQ. The findings of Van Luit and Naglieri (1999) therefore suggest that strategy instruction is an effective means of intervening to improve student performance with solving multiplication and division problems.

To summarize, different mathematics interventions have been successful in improving early numeracy (Van Luit & Schopman, 2000) and calculation skills (Mattingly & Bott, 1990; Miller &

Mercer, 1993; Van Luit & Naglieri, 1999) in children with MIDs and children with LDs; while the drill-and-practice intervention employed in Podell et al. (1992) was not. Notably, the studies by Van Luit and colleagues were the only ones that employed random assignment to study conditions *and* were effective in improving mathematics skills. Further, Van Luit and colleagues demonstrated that early numeracy and arithmetic problem solving skills are amenable to interventions explicitly designed to target specific areas of mathematics development. In particular, strategy instruction was effective in improving multiplication and division problem solving accuracy (Van Luit & Naglieri, 1992), while an early numeracy program that utilized a concrete-semiconcrete-abstract instructional sequence was effective in improving counting related skills (Van Luit & Schopman, 2000). In contrast to these positive findings, utilization of a CAI program, Math Blaster, which employs a drill-and-practice format (Podell et al., 1992), was not an effective intervention method for improving accuracy in solving basic arithmetic skills in children with mild disabilities.

1.6 Overview of the Present Study

An inference drawn from early work (Parmar et al., 1994) was that the nature of mathematics achievement in students with MIDs and students with LDs was different. Despite this potential difference, students from both special populations have responded well to interventions that have targeted early numeracy (Van Luit & Schopman, 2000) and arithmetic calculation (Mattingly & Bott, 1990; Miller & Mercer, 1993; Van Luit & Naglieri, 1992) skills. Although the discussed studies have contributed to our understanding of the nature of mathematics achievement and the influence of intervention as it relates to each special population, the aforementioned research is limited in important ways.

With respect to the nature of mathematics achievement, Parmar et al. (1994) suggested that mathematics achievement between children with MMR and children with LDs is substantially different; that qualitative and quantitative differences in mathematics skills are a consequence of between group IQ score differences; and that as a result, differentiating mathematics intervention (or instruction) according to this student characteristic should be considered. Unfortunately, Parmar et al. (1994) did not have access to student files and consequently, their IQ scores. Accounting for IQ within the statistical analyses, may have shown that the performance of children in both groups were more closely approximated to one another than the researchers concluded (see Jordan, Hanich, & Kaplan, 2003). Subsequently, the absence of IQ score data threatens the validity of the aforementioned conclusion and leaves the question concerning the nature of mathematics achievement is relates to children with MIDs and children with LDs unresolved. The present study will therefore systematically evaluate the nature (or structure) of mathematics achievement in each group of children.

In regard to students' response to mathematics intervention, empirical work has failed to address the possibility that measurement of mathematics achievement in children with MIDs and children with LDs was biased (or unreliable). Specifically, studies that employed group design methods (i.e., Podell et al., 1992; Van Luit & Naglieri, 1999; Van Luit & Schopman, 2000) assumed, perhaps fallaciously, that the nature and measurement of mathematics achievement was equivalent between children with MIDs and children with LDs at pre- and post-intervention time points. Failure to establish measurement equivalence/invariance (ME/I) between groups across time points leaves the possibility of differential item (or subtest) functioning unanswered. Thus, any given mathematics measure (or subtest) may have been biased such that children from one of the special populations (or intervention groups) responded to different attributes of the measures compared to children from the other special population. Moreover, the mathematics intervention may have altered the structure of mathematics achievement as measured over time (Vandenberg & Lance, 2000) in one of the groups (i.e., special population or intervention condition). As a result, systematic inaccuracies or variability in the information provided (e.g., measurement non-invariance due to variable language demands across items or subtests) may have biased the results concerning between group differences and/or students' response to the mathematics intervention (Brown, 2006).

In evaluating mathematics achievement growth within the employed studies, one study (Parmar et al., 1994) concluded that students with LDs evidenced greater growth rates than their age-equivalent peers with MMR. The use of a cross-sectional research design as employed in Parmar et al. (1994) is inadequate to demonstrate developmental trends or change over time due to the lack of time precedence, which requires a longitudinal study design (Whitley, 2002). Therefore, the conclusion that students with LDs evidence greater growth rates than their age-equivalent peers with MMR is therefore not supported by the methodology employed.

In other studies (i.e., Podell et al., 1992; Van Luit & Naglieri, 1992; Van Luit & Schopman, 2000), a pre-, post-test design was employed that utilized analysis of variance (ANOVA). Shortcomings associated with traditional methods such as ANOVA include (a) assuming that change in the conceptual domain is linear; (b) assuming that measurement of the conceptual domain is equivalent between groups across time, which is especially problematic when groups are composed of children from separate special populations; and that (c) tests of mean differences are not corrected for measurement error (Brown, 2006; Vandenberg & Lance, 2000). Acceleration in average growth for mathematical skills may be curvilinear as opposed to linear. Subsequently, forcing curvilinear data to fit a linear model of change (as in ANOVA) results in specification error and can influence the stability of the parameter estimates (e.g., over or underestimate estimates). In contrast to the use of repeated measures ANOVA, the present study will utilize growth curve modeling with repeated measures within the structural equation modeling (SEM) framework. It is advantageous to utilize latent growth curve modeling when possible. For instance, with data from three or more time points, a sample's average level for a given mathematics outcome (or competency) at each time point and their average rate of growth over time, can be estimated using growth curve modeling (Raudenbush & Bryk, 2002). Further, growth curve modeling is flexible such that measurement time points do not need to be equally distributed. With respect to SEM, using confirmatory factor analysis (CFA) is advantageous because it partitions the conceptual construct (e.g., mathematics achievement) into true score unique variance and random error variance (e.g., measurement error). Consequently, the biasing effects of random measurement errors can be accounted for (Medsker, Williams, & Holahan, 1994) and the distorting effects of measurement error on parameter estimates are mitigated (Chan, 1998; Vandenberg & Lance, 2000). In short, in comparison to using manifest indicators in ANOVA, using latent variables in SEM remedies problems related to poor measurement reliability and consequently, measurement error (Kline, 2011). Thus, growth curve modeling in an SEM framework provides a more flexible and stronger analytical methodology for investigating change over time compared to more traditional methods such as repeated-measures ANOVA.

A final limitation of the early empirical work cited, was that students with LDs were not further characterized according to their SLD status (i.e., RD, MD, MDRD). Failure to differentiate groups according to their SLD may have masked substantial group differences (e.g. Fuchs et al., 2004; Geary, Hamson, & Hoard, 2000; Geary et al., 1999; Jordan & Hanich, 2000). For instance, empirical evidence shows that students with MDRD evidence greater difficulties than students with MD on number comprehension tasks (Geary et al., 2000; 1999) and untimed arithmetic calculations (Hanich, Jordan, Kaplan, & Dick, 2001). Moreover, not further characterizing children with LD according to their specific status (i.e., SLD), prevents investigation of the relationship between SLD status and intervention outcome. Consequently, substantial group differences concerned with students' response to mathematics intervention may have been masked. In the present study, participants include students identified as mildly intellectually disabled and students with a SLD, RD.

With respect to remediating mathematics difficulties, historically, poor instruction has been cited as a primary cause of mathematics difficulties in students with disabilities (e.g., Carnine, 1991; Cawley, Fitzmaurice-Hayes, & Shaw, 1988; Cawley, Miller, & School, 1987; Kelly, Gersten, & Carnine, 1990; Miller & Mercer, 1993; 1997; Van Luit & Schopman, 2000; Wilson & Sindelar, 1991). While the potential exists that students with MIDs and students with LDs respond differentially to intervention, little research has examined this possibility. Of the few studies that included children from both special populations, the results suggest that the use of concrete, semiconcrete, and abstract instructional materials can improve early numeracy skills (Van Luit & Schopman, 2000) and can increase student's proficiency with solving addition, subtraction, and coin sum problems (Miller & Mercer, 1993). Further, time-delay procedures where the interval of time between the teacher's presentation of task directions paired with a novel stimulus and a controlling prompt, or the teacher's model of the correct response have been employed to improve student's multiplication facts proficiency (Mattingly & Bott, 1990); and students with MMR and LD have been shown to benefit from strategy instruction related to solving multiplication and division problems (Van Luit & Naglieri, 1999). However, some of these studies (Mattingly & Bott, 1990; Miller & Mercer, 1993) relied on small sample sizes and single-subject research methodology, which is characterized by high internal validity but poor external validity. Further, descriptions concerning one of the most debated characteristics of defining disability, IQ (see Ferrari, 2009; Fletcher, Lyon, Fuchs, & Barnes, 2007; Schalock, 2011), were not uniformly or clearly described in the reviewed literature as it related to subgroups of study participants.

Consequently, a need exists for a longitudinal study that systematically examines the nature (or structure) of mathematics achievement in elementary aged children with MIDs and their peers with LDs. Additionally, when two potentially separate groups of children are included within the same empirical study and analyses, ME/I should be examined between groups over time. Finally, structural equation modeling and latent growth curves should be used in place of ANOVA when possible. For the present study, participant data from two separate, completed randomized control trials with elementary aged students was utilized. One study included students with MIDs (n = 265), while the other included students with RD (n = 137). In each study, one group received a mathematics intervention, and the comparison group, intensive reading intervention. In both studies, mathematics achievement data were collected at three consecutive time points during the course of the school year. Measurement equivalence/invariance was evaluated, followed by an examination of the students' response to an evidence-based mathematics intervention.

1.7 Research Aims

Given the growing understanding that mathematics achievement is related to educational and occupational success, the paucity of empirical studies devoted to growing this field of research is unfortunate. Moreover, the majority of the research in this area has been concerned with mathematics development and difficulties as they relate to children with mathematics learning disabilities (MD) (Geary, 2013). In doing so, the study of mathematics development as it relates to children with MIDs has been neglected. Consequently, very little is known about mathematics achievement in children with MIDs (Brankaer, Ghesquière, & De Smedt, 2011; Foster, Sevcik, Romski, & Morris, 2014). Also, whereas the majority of studies concerned with mathematics development have focused on arithmetic calculations (Fuchs et al., 2004; Gersten et al., 2005), the present study will investigate mathematics achievement more broadly. One aim of the present study is to therefore investigate the nature (or structure) of mathematics achievement conceptualized as proficiency in skills related to the following areas: numeration, geometry, addition, subtraction, measurement, and time/money; areas of achievement that map onto the CCSSM (2010) standards.

Early work concerned with the nature of mathematics achievement and response to intervention as it relates to children with MIDs has been limited. Therefore, a rigorous longitudinal study that includes children with MID is needed. The present study systematically examines the nature (or structure) of mathematics achievement and students response to a mathematics intervention in a relatively large sample of elementary aged students' with MID and students' with RD. Of children with SLDs, those with RD, by definition, show an advantage in mathematical skill development over their peers with MD and MDRD, and more closely approximates that of typically achieving children (see Fuchs et al., 1994; Geary, Hamson, & Hoard, 2000; Geary, Hoard, Hamson, 1999; Jordan, Hanich, & Kaplan, 2003; Jordan, Kaplan, & Hanich, 2002). Inclusion of students with RDs in the present study can therefore enable inferences concerning the acquisition of mathematics proficiency in children with MIDs as it relates to typically achieving children. In particular, finding that the structure of mathematics achievement is equivalent in children with MIDs and those with RDs suggests that children from both special populations follow similar, if not the same, sequence in developing mathematics proficiency as typically achieving children. Subsequently, this provides a basis that can help researchers understand observed growth and change in its ordered and sequential manner.

In order to draw clear inferences concerning students' response to intervention, limitations of early research are addressed. In particular, the nature (or structure) of mathematics achievement as it relates to children with MIDs and children with RDs will be investigated through CFA; after which, measurement of mathematics achievement will be systematically examined through ME/I evaluation. Establishing longitudinal ME/I is a necessary precondition to studying longitudinal change. Without first establishing longitudinal measurement stability, it cannot be determined that temporal change observed in a construct (i.e., mathematics achievement) is due to true change or changes related to precision in measurement of the construct, or changes in the construct itself that varies across time (Brown, 2006).

For between group studies involving different populations, establishing ME/I between groups is a necessary precondition to making meaningful inferences concerned with mean group differences. This is because it is necessary to rule out the potential for differential item (or subtest) functioning such that group differences are not a consequence of the target groups responding to different attributes of an item (e.g., expressive language skills necessary to correctly respond to an item on a norm-referenced mathematical test). Thus, in the absence of ME/I, it is misleading to analyze and interpret longitudinal change and/or group mean differences. Therefore, whereas early research (Parmar et al., 1994; Podell et al., 1992; Van Luit & Naglieri, 1999; Van Luit & Schopman, 2000) implicitly assumed that the structure and measurement of mathematics achievement was equivalent, the present study will explicitly evaluate the tenability of this assumption through ME/I evaluation. In particular, students will be combined in a single group for ME/I evaluation; however, in the case that measurement non-invariance is identified, students will be separated according to intervention condition assignment because intervention effects can result in non-invariance of model parameters (McArdle, 1996).

Following ME/I evaluation, students' response to an evidence-based mathematics intervention is assessed using the SEM framework. In doing so, the present study examines, whether or not, children with MIDs and children RDs benefit from a mathematics intervention as evidenced by change over time captured by a norm-referenced test. Given present educational policy (Individuals with Disabilities Education Act, 2004) that expects students with disabilities to make progress in their mathematics curricula and to demonstrate proficiency on high stakes testing (e.g., Georgia High School Graduation Exam), it is important to identify interventions that educators can implement to remediate mathematics difficulties in children with mild disabilities.

1.8 Research Questions

In order to examine the structure of mathematics achievement, its measurement, and students' response to mathematics intervention, the following three research questions were addressed.

Question 1. Is the nature (or structure) of mathematics achievement in children with MIDs and children with RDs temporally stable? That is, is the form (or configuration) of mathematics achievement equivalent over time in elementary aged children with MIDs and others with RDs? Although children from these groups represent potentially separate special populations, prior work (Foster et al., 2014; Wise et al., 2008) has demonstrated that children with MID and those with RD evidence the same types of reading and mathematics relationships. It is expected that the form of mathematics achievement will be equivalent between groups across time. In other words, it is expected that the structure (i.e., the number of factors and pattern of factor-indicator loadings) of this conceptual domain will be stable between groups across pre-, mid-, and post-intervention time points.

Question 2. Is the measurement of mathematics achievement in children with MIDs and children with RDs temporally equivalent? That is, does the measurement of mathematics achievement function equivalently over time in elementary aged children with MIDs and others with RDs? It is expected that the measurement of mathematics achievement is equivalent in children from these special populations as evidenced by equal form (see *Question 1*) and at minimum, partially equivalent factor loadings and intercepts between groups across the three time points.

Question 3. Did elementary aged students' with MIDs and students with RDs who participated in an evidenced-based mathematics intervention show increased mathematics achievement growth compared to their peers who participated in a reading intervention? The present study is the first systematic attempt to evaluate students with MIDs and students with RDs response to intervention following ME/I evaluation. Early research (Mattingly & Bott, 1990; Miller & Mercer, 1993; Van Luit & Naglieri, 1992; Van Luit & Schopman, 2000), however, suggests that interventions that have targeted early numeracy and arithmetic skills have been successful in improving mathematics proficiency. It is therefore expected that students who participated in the mathematics intervention will evidence more growth in mathematics than students that participated in a reading intervention.

2. METHOD

The data analyzed for this study were collected as part of two completed reading intervention efficacy projects for elementary aged children that differed according to disability status (Sevcik, 2005 and Morris, 1996). Sevcik (2005) collected data over the course of five school years from August 2005 to May 2010 and focused on second to fifth grade students diagnosed by their local school districts with MID. Morris (1996) collected data over the course of five years from May 1996 to May 2001 and focused on second and third grade students diagnosed with RD. Due to the focus (i.e., reading intervention) of Sevcik (2005) and Morris (1996), participants in both projects had the opportunity to be randomly assigned to a reading condition each of the five years, whereas the opportunity to be randomly assigned to the mathematics condition was only available in years one through four. Further, the data analyzed in this study are from three time points (pre-intervention, intervention midpoint, and post-intervention);

2.1 Participants

Children with Mild Intellectual Disabilities

Participants with MIDs were screened with a set of inclusionary and exclusionary criteria. Inclusionary criteria included measured IQ from 50-70 and poor or no reading skills (below the 10th percentile on standardized reading measures). Participants were excluded if they did not speak English, had a history of hearing impairment (<25 dB at 500+Hz bilaterally), a history of uncorrected visual impairment (<20/40), and/or had serious emotional/psychiatric disturbance (e.g., major depression, psychosis) as described in parent reports.

The 265 participants with MID were assessed by and met their local school district's eligibility criteria for MID. IQ scores were obtained from each child's school when available. Student M_{IQ} (*sd*) = 63.03 (9.64). Etiology of the intellectual impairments was heterogeneous and included Down syndrome, Fragile X syndrome, and etiology unknown. Of the participants, 96 (36.2%) were girls and 169 (63.8%) were boys. In regard to racial and ethnic diversity, there were 6 (2.3%) Asian, 150 (56.6%) African American, 43 (16.2%) Hispanic, 53 (20.0%) Caucasian, and 12 (4.5%) Multi-racial students (race was not reported for one participant). Sample M_{age} (*sd*) = 9.27 (1.34) years and ranged from 6.67 to 12.25 years.

Finally, participants with MID were close to equally distributed across grade levels with 84 (31.7%) second grade, 58 (21.9%) third grade, 69 (26.0%) fourth grade, and 54 (20.4%) fifth grade students. Of these 265 participants, 182 were randomized to a reading intervention, and 83 were randomized to a mathematics intervention.

Children with Reading Disabilities

Participants with RDs were screened with a set of inclusionary and exclusionary criteria. Inclusionary criteria included the low achievement (LA) and/or Ability-Achievement Regression Corrected Discrepancy (DISC) definitions for RD. Participants with a *Kaufman-Brief Intelligence Test* (K-BIT; Kaufman & Kaufman, 1990) composite score greater than 70 and whose reading skills were equal to, or less than, a standard deviation score of 85 on the *Woodcock Reading Mastery Test-R* (WRMT-R; Woodcock, 1987) were identified as meeting the LA criteria for reading disability. Participants whose reading performance was at least one standard error of the estimate below their Expected Achievement Standard Score (EASS), calculated based on an average correlation of 0.60 between measures of reading performance and intellectual ability, were included under the DISC criteria. As in Sevcik (2005), participants were excluded if they did not speak English, had a history of hearing impairment (<25 dB at 500+Hz bilaterally), a history of uncorrected visual impairment (<20/40), and/or had serious emotional/psychiatric disturbance (e.g., major depression, psychosis) as described in parent reports. Additionally, children were excluded if they had repeated a grade or had a K-BIT Composite Score below 70. Participants who had repeated a grade were excluded in attempt to control for the amount of previous educational experience of the children.

Of the 279 participants in Morris (1996), mathematics achievement was measured in 137. These 137 participants, all with RD were from three large metropolitan areas (Atlanta: n = 47 [34.3%], Boston: n = 29 [21.2%] and Toronto: n = 61 [44.5%]). In contrast to Sevcik (2005), all students were independently evaluated for RDs. Mean reading achievement measured by the WRMT was 77.33 (sd = 11.97) and mean IQ as measured by the K-BIT was 91.09 (sd = 11.04). Of the participants, 47 (34.3%) were girls and 90 (65.7%) were boys. In regard to racial and ethnic diversity, there were 67 (48.9%)

African American and 70 (51.1%) Caucasian students. Sample M_{age} (*sd*) = 7.51 (0.56) years and ranged from 6.42 to 8.83 years. Finally, with respect to grade level 102 (74.5%) second grade and 35 (25.5%) third grade students with RDs are represented in the present analyses. Of these 137 participants, 70 were randomized to a reading intervention and 65 were randomized to a mathematics intervention.

2.2 Assessment Instruments

The *KeyMath-Revised Diagnostic Inventory* (Connolly, 1988) was administered as part of a larger assessment battery. This norm-referenced mathematics measure was selected because it is widely used in educational and remedial outcome research, psychometrically appropriate for growth curve modeling, and because it has adequate reliabilities and validity. Finally, the KeyMath-R will allow for comparison of the sample's mathematics achievement and abilities with those from other published empirical studies.

Students in Sevcik (2005) and Morris (1996) were evaluated throughout the school year and the data in this study are from three time points (prior to random assignment to a study condition, at the intervention mid-point, and following the completion of the intervention). The number of intervention hours differed between the studies. Students in Sevcik (2005) received up to 120 hours of intervention with mid-point assessment occurring after 60 hours. In contrast, students in Morris (1996) received up to 70 hours of intervention with mid-point assessment occurring after 35 hours. The present study will analyze raw rather than standard scores because the KeyMath-R examiner's manual does not report including children with disabilities in the norming standardization procedures. Therefore, using standard scores would likely underestimate student performance and restrict variability in scores due to measurement sensitivity issues (i.e., floor effects) in the data and consequently, result in incorrect parameter estimates that could mask the true relationships between the mathematics indicators.

Measures of Mathematics

Students' mathematics achievement was measured using six subtests from the KeyMath-R: Numeration, Geometry, Addition, Subtraction, Measurement, and Time/Money. The KeyMath-R is widely used in education and research settings and evidences sufficient reliability. For children between 6 and 12 years of age, split-half reliability coefficients corresponding to each subtest were generally stronger for Spring (*r* range: between .66—Addition, and .92—Time/Money) than Fall (*r* range: between .57—Measurement, and .93—Time/Money). The domain-referenced scope and sequence of the KeyMath-R identified hierarchies of concepts and skills. The subtests above are divided into the three following areas.

Basic Concepts Measures. The Numeration subtest measures students' understanding of quantity, order, and place value; whereas the Geometry subtest measures their understanding of spatial and attribute relations, two-dimensional shapes, coordinates and transformations, and three-dimensional shapes.

Operations Measures. The Addition and Subtraction subtests assess students' understanding of arithmetic facts, algorithms to add/subtract whole numbers, and adding/subtracting rational numbers. Written calculation begins with item seven on each subtest, respectively.

Application Measures. The Measurement subtest evaluates students' understanding of comparisons using standard and non-standard units related to length, area, weight, and capacity. The Time/Money subtest measures identification of passage of time, use of clocks and clock units, and understanding monetary amounts from one dollar to one hundred dollars and business transactions.

2.3 Mathematics and Reading Intervention Programs

Mathematics Intervention

Both of the completed larger projects utilized the same direct instruction mathematics programs, Distar Arithmetic II (Engelmann & Carnine, 1976) and/or Connecting Math Concepts (CMC; Engelmann & Carnine, 1992), which was a function student's curriculum-based placement testing results. Both of the larger projects also utilized Base Ten Blocks (McLean, Laycock, & Smart, 1990) as a supplement to the direct instruction mathematics program(s). Distar II and CMC are direct instruction programs with lessons organized around multiple concepts and skills, each of which is addressed for only 5 to 10 minutes in a given day and then revisited day-after-day for many lessons. In both mathematics programs, students are explicitly taught concepts and strategies for solving arithmetic computations and word problems. Distar II was employed with students who demonstrated a need to develop and build prerequisite and basic skills such as rote counting, numeral copying, symbol identification, and basic addition, subtraction and place value skills. In contrast, CMC was used with more advanced students. Instruction quickly advances from counting activities and symbol identification to learning concepts such as equality, discriminating between differing numerical magnitudes, and understanding number relationships and using the number line, to solving arithmetic and application problems involving money, measurement, and estimation as well as solving problems involving fractions and word problems. More advanced students in the CMC curriculum series were also explicitly taught skills related to geometry (e.g., identifying shapes, computing perimeter and area), and analyzing data presented in tables and interpreting graphs.

Base Ten Blocks was used as a supplement to the interventions described above. Corresponding activities taught computational procedures in concrete format to help students consolidate numeration, number line, and arithmetic concepts. Additionally, Morris's (1996) mathematics intervention condition included a component focused on teaching students to listen for critical words and implementing a four-step metacognitive strategy (*think*, *plan*, *do*, *check*) when solving word problems. Note that although the KeyMath-R includes a word problem-solving subtest, it was not utilized in the present study.

Reading Intervention

Both of the completed parent projects used the same reading programs: Phonological Analysis and Blending/Direct Instruction Program (PHAB/DI) and the Retrieval-Rate, Accuracy, Vocabulary Elaboration and Orthography Program (RAVE-O). However, in Morris (1996), measurement of mathematics achievement was limited to children randomly assigned to mathematics and the PHAB/DI conditions. In Sevcik (2005) mathematics achievement was measured in all children regardless of their intervention condition assignment. PHAB/DI trains children in phonological analysis and blending skills in the context of printed presentations and direct instruction of letter-sound and letter-cluster-sound correspondences; whereas RAVE-O was designed to add to a phonological foundation in reading instruction and emphasizes meaning, rapid retrieval in oral and written language, and efficient orthographic decoding. Finally, Morris (1996) paired the Classroom Survival Skills Program (CSS) with each intervention condition. CSS is not theoretically informed and trains students in classroom etiquette, life skills, and organizational strategies, with an emphasis on academic problem solving and self-help techniques.

Teacher Training and Treatment Integrity

Teachers that led the instructional groups were employees of Georgia State University and were certified to teach in the state of Georgia. All teachers received intensive training (three to seven days) in delivering intervention components within the respective research projects. Additionally, weekly meetings were used to provide ongoing instructional support to teachers and the use of an observational rating form documented treatment integrity. Ten percent of the total number of instructional sessions also were videotaped and indicated that the intervention programs were being carried out as planned. Finally, daily logs of the sessions were kept and reviewed weekly to provide close monitoring of instructional issues as they arose.

Statistical Analyses and Power

The present study utilizes analytic procedures within the SEM framework. In regard to sample size and consequently, power, several rules of thumb have been proposed. For instance, in order to avoid model nonconvergence and inadmissible solutions (e.g., negative variance estimates), Boosma and Hoogland (2001) recommend a minimum sample size (*N*) of 200 and that the ratio of number of indicators per factor equal 3:1 or 4:1 (given *N* of 200). The present sample consists of data for 402 total participants (MIDs, n = 265; RDs, n = 137) and the corresponding ratio of indicators per factor (6:1) exceed the rules of thumb proposed by Boosma and Hoogland (2001). In regard to model fit of specific analyses, recommendations provided by Bentler (2007) will be followed. As such, model fit will be evaluated in terms of the χ^2 test of exact fit as well as the comparative fit index (CFI; Bentler, 1990), root mean square error of approximation (RMSEA; Steiger, 1990), and standardized root mean square residual (SRMR).

2.4 Procedure

In both larger projects (Sevcik, 2005; Morris, 1996) the procedures were similar. School administrators and teachers initially identified children who met the state and local school district’s criteria for the respective disability. Packets that contained the study's description, a consent form, and a demographic survey were sent home with identified students. After students returned signed consent forms and provided assent, they were administered an assessment battery by trained project personnel (graduate students and faculty). Test administration occurred within the student's local school during the typical school day and required between three and five hours over the course of a few days.

Following pre-intervention assessment in Sevcik (2005), small groups of children and teachers were randomly assigned to one of three study conditions (two reading interventions or a mathematics intervention). Groups of four to five children, on average, were taught by trained certified teachers for up to 120 instructional hours during a school year. All children were evaluated at the beginning (0 hours), middle (60 hours), and end of the intervention (up to 120 hours).

Following pre-intervention measurement testing in Morris (1996), small groups of children and teachers were randomly assigned to one of four study conditions (three reading interventions or a mathematic intervention); although, only two intervention conditions are relevant to the present study (PHAB/DI and Mathematics). Groups of four to five children, on average, were taught by trained certified teachers for up to 70 instructional hours during the school year. All children were evaluated at the beginning (0 hours), middle (35 hours), and end of the intervention (70 hours).

To ensure accuracy and quality control of data, all data were entered into SPSS 18 using a double entry procedure with two independently working researchers. Crosschecks between the two entries were run to determine potential inconsistencies. If an inconsistency was found, the original test protocol was referenced, the data corrected, and cross checks run again. This process was continued for all data until no inconsistencies were found.

3. RESULTS

3.1 Descriptive Statistics

In order to evaluate the nature (or structure) of mathematics achievement and students' response to a mathematics intervention, data from students with MIDs and students with RDs that participated in the two larger projects were analyzed. Descriptive statistics according to intervention condition are presented in Tables 1, 2, and 3 for the pre-, mid-, and post-intervention time points, respectively (see Appendices 1 and 2 for descriptive statistics differentiated by student disability status, and in the combined group of students, respectively). In each table, means for each variable represent the average number of items correct for a given subtest. Examination of the distributions for several of the KeyMath-R subtests differentiated by intervention condition type (mathematics and reading) indicated that they evidenced non-normal distributions. For instance, skew and kurtosis statistics can be converted to *z* scores by subtracting the mean of the respective distribution (in this case 0) from the target score and then dividing by the standard error (SE_{skewness}; SE_{kurtosis}) of the distribution. An absolute value greater than 1.96 is significant a $p < .05$, whereas an absolute value of 2.58 is significant a $p < .01$ (Field, 2012). For the pre-intervention data (Table 1), by dividing the skewness statistic by its standard error resulted in a value greater than 1.96 for five (Numeration, Addition, Subtraction, Measurement, and Time/Money) of the six subtests for the mathematics intervention group and all of subtests in the reading intervention group; whereas, dividing the kurtosis statistic by its standard error indicated significant ($p < .05$) kurtosis for four (Numeration, Geometry, Measurement, and Time/Money) of the six subtests in the mathematics group and three of the subtests (Geometry, Subtraction, and Time/Money) in the reading group. Further investigation of non-normality was examined visually through histograms and *q*-*q* plots of the data. Examination of histograms corresponding to each subtest differentiated by intervention condition indicated that three (Subtraction, Measurement, and Time/Money) of the six subtests evidenced positive skew across both intervention groups, which is in part due to floor effects. Examination of *q*-*q* plots confirmed the previous findings. Finally, investigation of significant skewness and kurtosis at the second and third time points, also suggested that the data were significantly ($p < .05$) skewed and/or kurtotic. As displayed in Tables 1, 2, and 3, the number of participants that earned a score of zero on the respective subtests varied within and across each time point. In short, most of the indicator's distributions are characterized by non-normality. Methods for addressing issues related to distribution non-normality and floor effects will be discussed in the data analysis section.

Table 1. Descriptive Statistics by Intervention Group Across Studies: Pre-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
Math Group									
NUM	149	5.94	3.21	5.0	0	19	1	1.01 (.20)	1.35 (.40)
GEO	149	5.48	3.93	5.0	0	14	17	0.36 (.20)	-0.82 (.40)
ADD	149	4.17	2.72	4.0	0	12	9	0.53 (.20)	-0.33 (.40)
SUB	149	1.95	1.98	1.0	0	10	45	0.99 (.20)	0.32 (.40)
MST	149	4.50	3.40	3.0	0	12	10	0.57 (.20)	-0.94 (.40)
TIMO	149	2.47	2.37	2.0	0	10	36	1.14 (.20)	1.09 (.40)
Reading Group									
NUM	248	6.20	3.21	6.0	0	18	2	0.68 (.16)	0.10 (.31)
GEO	247	5.08	3.78	5.0	0	15	35	0.36 (.16)	-0.72 (.31)
ADD	248	4.13	3.14	3.5	0	14	25	0.70 (.16)	-0.15 (.31)
SUB	248	2.01	2.05	1.0	0	10	71	1.09 (.16)	0.80 (.31)
MST	247	4.02	3.29	3.0	0	13	29	0.82 (.16)	-0.31 (.31)
TIMO	247	2.97	2.57	2.0	0	12	36	1.16 (.16)	1.32 (.31)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Table 2. Descriptive Statistics by Intervention Group Across Studies: Mid-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
Math Group									
NUM	143	7.55	3.70	7.0	0	18	1	0.51 (.20)	-0.51 (.40)
GEO	144	6.82	4.49	7.0	0	16	11	0.15 (.20)	-1.03 (.40)
ADD	144	5.30	3.26	5.0	0	12	9	0.18 (.20)	-1.04 (.40)
SUB	144	2.78	2.48	2.0	0	11	28	0.89 (.20)	0.23 (.40)
MST	144	5.13	3.92	4.0	0	13	12	0.31 (.20)	-1.30 (.40)
TIMO	144	3.62	2.88	3.0	0	15	15	1.05 (.20)	1.31 (.40)
Reading Group									
NUM	242	7.18	3.59	6.0	1	19	0	0.59 (.16)	-0.36 (.31)
GEO	241	6.43	4.07	6.0	0	16	20	0.17 (.16)	-0.82 (.31)
ADD	242	5.43	3.32	5.0	0	14	10	0.30 (.16)	-0.69 (.31)
SUB	241	2.43	2.40	2.0	0	10	62	1.00 (.16)	0.30 (.31)
MST	241	4.81	3.46	3.0	0	16	13	0.60 (.16)	-0.59 (.31)
TIMO	241	3.80	2.96	3.0	0	16	21	1.07 (.16)	1.53 (.31)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Table 3. Descriptive Statistics by Intervention Group Across Studies: Post-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis(SE)*
Math Group									
NUM	143	8.25	3.95	7.0	0	20	1	0.63 (.20)	-0.11 (.40)
GEO	143	7.41	4.26	7.0	0	16	8	0.10 (.20)	-0.78 (.40)
ADD	143	6.05	3.49	6.0	0	14	8	0.63 (.20)	-0.90 (.40)
SUB	143	3.17	2.76	3.0	0	11	27	0.81 (.20)	0.23(.40)
MST	143	5.70	3.92	5.0	0	15	8	0.44 (.20)	-0.95 (.40)
TIMO	143	4.66	3.82	4.0	0	21	11	1.53 (.20)	3.25 (.40)
Reading Group									
NUM	238	7.87	3.72	7.0	1	18	0	0.35 (.16)	-0.73 (.31)
GEO	238	7.39	4.13	7.0	0	17	10	0.13 (.16)	-0.71 (.31)
ADD	238	6.16	3.53	6.0	0	14	7	0.16 (.16)	-0.79 (.31)
SUB	238	3.01	2.50	3.0	0	10	37	0.72 (.16)	-0.20 (.31)
MST	238	5.41	3.76	4.5	0	16	15	0.38 (.16)	-0.82 (.31)
TIMO	238	4.49	3.19	4.0	0	16	16	0.90 (.16)	0.84 (.31)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

3.2 Missing Data Patterns

Instances of missing data were investigated by hand and through M*plus* software (Muthén & Muthén, 1998-2010), which identified 11 patterns of missing data. In general, missingness was assumed to be at random (MAR). The first pattern was no missing data (*n* = 367). The second through seventh patterns of missing data each represented one participant who was missing scores for the pre-intervention Measurement and Time/Money subtests, pre-intervention Geometry subtest, all six subtests at pre-intervention, mid-point Measurement and Time/Money subtests, mid-point Subtraction subtest, and all subtests at mid-point except Time/Money, respectively. The eighth pattern of missing data represented participants who were missing scores for all pre-intervention and mid-point KeyMath-R subtests (*n* = 4). The ninth pattern represented participants who were missing scores for the Time/Money subtest administered at the post-intervention time point (*n* = 2). The tenth pattern represented participants who were missing data for all KeyMath-R subtests administered at the post-intervention assessment (*n* = 10). This group of participants represented those who left their respective study early. Finally, the eleventh pattern represented participants who had missing scores across all measures for the mid-point and post-intervention time points (*n* = 11). This group's missingness also was due to leaving their respective study early. In summary, the largest source of missing data was due to participants leaving the study before its completion, respectively (total, *n* = 19). Nevertheless, 92.75% of participants had complete data for all of the measures administered across the three time points. Strategies for addressing issues related to missing data will be discussed in the analysis section.

3.3 Confirmatory Factor Analysis (CFA) for Mathematics Achievement

The nature (or structure) of mathematics achievement in children with MIDs and children with RDs was assessed utilizing confirmatory factor analysis (CFA) and ME/I evaluation with M*plus* (version 7) software. CFA is a strategy whose purpose is to identify latent constructs (or factors) that account for variation and covariation among a set of indicators. All aspects of the factor model are pre-specified (e.g., the number of factors, the pattern of indicator-factor loadings, etc. (Brown, 2006). CFA utilizes maximum likelihood (ML) estimation to find parameter estimates that maximize the likelihood of observing the given data if it were collected from the same population again. Additionally, ML is a full information (FIML) estimation method (also referred too as direct ML), which is a preferred method for handling missing data (Allison, 2003; Schafer & Graham, 2002). FIML makes use of all of the available information, even cases with missing data, when estimating parameters (Brown, 2006).

3.4 Data Preparation and Special Considerations

Prior to investigating the nature of mathematics achievement and subsequently, students' response to a mathematics intervention, steps were taken to prepare the data. Using SPSS 20, two separate data sets, one for students with MIDs and another for students with RDs were merged. The merged data set included all variables reported in the present study. An advantage of employing a one-sample approach as in the present study was that correlated errors could be estimated and accounted for when estimating other model parameters (Brown, 2006). In addition, maximum likelihood estimation with robust standard errors (MLR; Bentler, 1995) was chosen for the present analyses in order to address non-normality of the distributions and non-independence of observations for the included subtests (see Tables 1, 2, and 3). An advantage of MLR estimation (χ^2_{SB}) is that it provides a correction for non-normality (e.g., floor effects) in continuous indicators (Yuan & Bentler, 2000). Thus, MLR estimation provides a model chi-square and standard errors of the parameter estimates that are corrected for non-normality.

3.5 Testing the Structure of Mathematics Achievement

The nature (or structure) of mathematics achievement was examined using ME/I evaluation within the context of CFA with nested χ^2 methods (i.e., difference testing). Evaluation of ME/I is a method that directly evaluates the tenability that a set of indicators (e.g., KeyMath-R subtests) reliably and validly assess a conceptual domain (e.g., mathematics achievement) between groups (multiple group CFA) and/or within groups across time (Curran & Hussong, 2009). Further, the use of nested models provides the opportunity to make direct statistical comparison of alternative models (or solutions) possible. Within this context, alternative (or subsequent) models are characterized by more constraints than the prior model and difference testing provides evidence that indicates whether or not the additional constraints significantly reduce model fit. As such, ME/I evaluation within the present study utilized the forward restriction method, which is recommended in Vandenberg and Lance (2000). The forward restriction method adds constraints to an unconstrained model such that, in the present study, ME/I evaluation proceeds from evaluation of equal form, to equal factor loadings, and finally, equal intercepts. The test of *equal form* evaluated the hypothesis that the same number of factors and pattern of factor-indicator loadings were temporally equivalent between groups of students differentiated by intervention condition participation. The test of *equal factor loadings* evaluated the hypothesis that the indicator's factor loadings were temporally equivalent between the two groups of children; and the test of *equal* *intercepts* evaluated the hypothesis that the intercept parameters were temporally equivalent between the groups.

Model fit was evaluated in terms of the Satorra-Bentler chi-square (χ^2_{SB}) test of model fit, the comparative fit index (CFI; Bentler, 1990), the root mean square error of approximation (RMSEA; Steiger, 1990), and the standardized root mean square residual (SRMR). When difference testing and the corresponding fit indices indicated acceptable model fit, the respective equality constraint remained in place and an additional equality constraint was included in the subsequent analysis. For instance, if the addition of the factor loadings equality constraint resulted in acceptable model fit, it remained in place while adding the intercept equality constraint within the subsequent analysis. Model fit of this new model, characterized by the additional specification for intercept equivalence, was then evaluated in comparison to the previous model (i.e., that did not include the intercept equality constraint). In cases where the inclusion of an additional equality constraint significantly reduced model fit, partial ME/I (see Byrne, Shavelson, & Muthén, 1989) was pursued. Returning to the previous example, if the addition of the intercept equality constraint resulted in significantly reduced model fit as indicated by difference testing (χ^2_{SBdiff}), analyses were carried out to determine if some, but not all, of the indicator intercepts, for example, were temporally equivalent between groups.

Questions 1 and 2. In order to address *Research Question 1* and examine the nature (or structure) of mathematics achievement in elementary aged students with MIDs and students with RDs, data corresponding to both groups were combined. After which, the structure of mathematics achievement was evaluated using the test of equal form. In order to address *Research Question 2* and examine the measurement of mathematics achievement, the test of equal form was followed by the test of equal factor loadings and then equal intercepts, respectively. In this ‘combined group’ context, each equality test was simultaneously employed across the three time points. In doing so, temporal stability in mathematics achievement was examined. Of the equality tests, it was expected that mathematics achievement would be characterized by equal form (or configural invariance) and equal factor loadings (or metric invariance).

However, because intervention effects can result in non-invariant model parameters, especially intercepts (McArdle, 1996), it was anticipated that the addition of the intercept equivalence (or scalar invariance) constraint (to the equal factor loadings model) would significantly reduce model fit as students in this combined group were randomly assigned to a reading (*n*, MID = 182, RD = 71; Total = 253) or mathematics (*n*, MID = 83, RD = 66; Total = 149) intervention condition.

For the test of equal form, each of the six KeyMath-R subtests were specified as indicators of a single mathematics achievement latent construct at each of the three time points (across population, study, and intervention group), respectively (see Figure 1). Accordingly, correlated residuals were specified to account for indicator specific method variance (or method effects) associated with repeated administrations of the same measure; however, for ease in interpretation, the specification of correlated residuals as well as the correlations between the latent mathematics achievement factors, are omitted below. Model identification was achieved by fixing the latent mathematics achievement means and variances at each of the three time points to zero. The overall fit for the equal form solution is presented in Table 4. Although, the χ^2_{SB} was significant, all other fit indices provided support for the hypothesis that the structure of mathematics achievement is unidimensional at each measurement occasion. That is, the same form (or configuration) is present at each time point in this combined group of children.

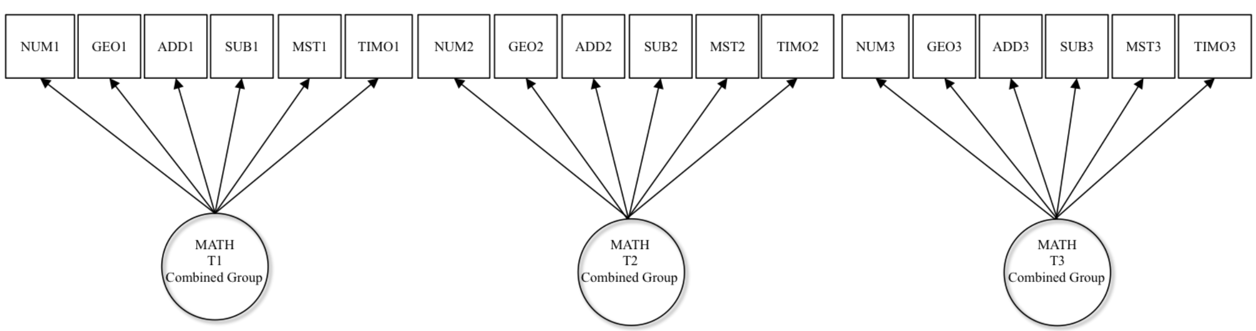

Figure 1. Combined Group Model: Equal Form. Specification of correlated residuals and correlations between latent mathematics achievement factors are omitted above. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. T1 = Pre-intervention time point, T2 = Mid-intervention time point, T3 = Post-intervention time point.

Table 4. Measurement Invariance Evaluation for Mathematics Achievement: Combined Group Model

	χ^2_{SB} (*df*)	χ^2_{SBdiff}	Δdf	CFI	RMSEA (90% CI)	$P_{RMSEA} \leq .05$	SRMR
Step 1. Equal Form	203.09 (114)***			.986	.044 (.034, .054)	.829	.024
Step 2. Equal Factor Loadings	221.35 (124)***	17.35	10	.985	.044 (.035, .054)	.836	.030
Step 3. Equal Intercepts	321.45 (134)***	100.10***	10	.970	.059 (.051, .067)	.035	.049

Note. χ^2_{SB} = Satorra-Bentler scaled χ^2; χ^2_{SBdiff} = difference test; Δdf = change in degrees of freedom; CFI = Comparative fit index, RMSEA = Root mean square error of approximation, $P_{RMSEA} \leq .05$ = Test of close fit, SRMR = Standardized root mean square residual.
$^{***}p < .001$.

In the subsequent analysis (Step 2) measurement stability was further evaluated by investigating whether or not the factor loadings were equivalent across time. Because ME/I evaluation focuses on the unstandardized relationships within the specified model, factor loadings are regression coefficients. That is, factor loadings represent the regression of the latent construct on the observed indicators. Building on the previous model specification (i.e., equal form with correlated residuals), the factor loadings for each of the six KeyMath-R subtests were simultaneously constrained to equality across the three time points (see Figure 2). In other words, the regression of each indicator on the latent construct was specified as equivalent across time. Model identification was achieved by fixing the latent means to zero; however, in contrast to the previous analysis, the metric of the latent variance was set by fixing the variance of mathematics achievement at the first time point to one, while freely estimating the latent variance at the latter two time points. The results of difference testing (see Table 4, Step 2) provided evidence for metric invariance. Table 4 displays the overall model fit as well as model fit statistics, which suggest that the equal factor loading model specification provides a good fit to the data. Moreover, the model χ^2_{SB} for the equal factor loadings solution was 221.35 (df = 124), which resulted in a non-significant χ^2 difference test, χ^2_{SBdiff} (10) = 17.35, p = 0.066. These findings therefore suggest that the factor loadings (i.e., measurement metric) are temporally stable. Thus, each of the six factor loadings were equivalent across time in this combined group model. In Figure 2, the estimates for the factor loadings are displayed *below* each indicator. As evident in Figure 2, the factor loadings are the same for each indicator at each of the three time points, which indicates that with each unit of change in the latent mathematics achievement factor, the Numeration subtest score, for example, is expected to change 2.87 units, the Geometry subtest, 2.85 units, the Addition subtest, 2.45 units, and so on.

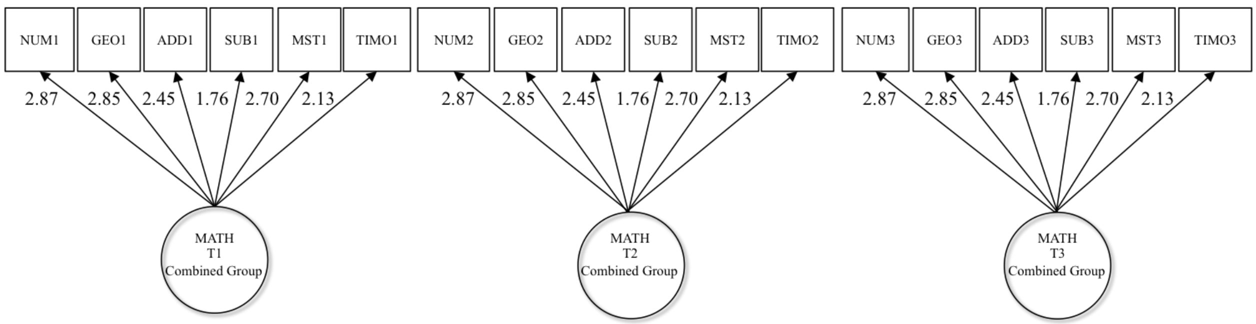

Figure 2. Combined Group Model: Equivalent Factor Loadings. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point. Factor loadings displayed below each indicator.

After demonstrating that the latent mathematics achievement construct was characterized by configural (Step 1) and metric (Step 2) invariance, equality of the intercepts was investigated (see Figure 3). As such, the intercepts for each of the six KeyMath-R subtests were constrained to equality across time while the two previous model constraints remained in place. Model identification was achieved by fixing the latent mathematics achievement mean and variance for the first time point to zero and one, respectively. In addition to the factor loadings, the estimates for intercept parameters are displayed *above* their indicator in Figure 3 and the asterisk indicates non-invariance. As displayed in Table 3, the model fit statistics are attenuated compared to the equal factor loadings model. Moreover, the model χ^2_{SB} of the equal intercepts solution was 321.45 (*df* = 134), which resulted in a significant difference test, χ^2_{SBdiff} (10) = 100.10, $p < 0.001$, suggesting that the equal intercepts solution fit significantly worse than the equal factor loadings solution specified in Step 2 (see Table 4).

Intercept parameters are interpreted as the model-implied origin of scale or where the mean would be given a level of the latent factor. Finding that the intercepts are not equal across the three time points and that by forcing them to equality, model fit suffers, suggest that the indicator's mean, changes significantly over time. Moreover, intercept non-invariance is evidence of instability of the scale of the latent mathematics achievement construct within the combined group context (identified in Figure 3 by the red intercept parameter with asterisk).

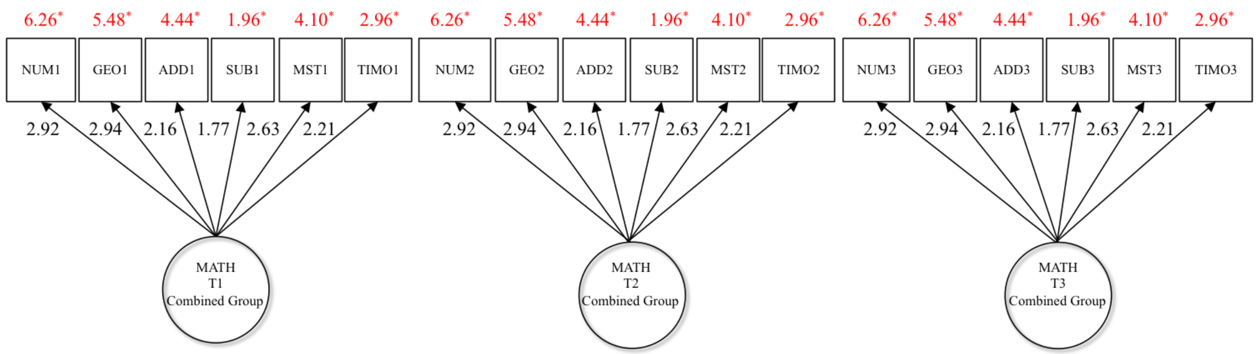

*Figure 3.*Combined Group Model: Equivalent Intercepts. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point. Factor loadings displayed *below* each indicator. Intercepts displayed *above* each indicator. Asterisk indicates non-invariance.

In summary, configural, metric, and scalar invariance were evaluated in a combined group model that included elementary school aged students with mild disabilities who participated in either a reading or mathematics intervention. It was demonstrated that a congeneric factor model provided the best representation of the underlying data structure at each of the three time points (i.e., equal form). Evidence of metric invariance (i.e., factor loading equivalence) also was established suggesting that the indicators evidenced comparable relationships to the latent mathematics achievement construct over time. In contrast, scalar invariance (i.e., intercept equivalence) could not be established implying that the indicator's location parameters (means) changed over time.

In responding to *Research Question 1*, evidence of configural invariance within the combined group model provides support for the hypothesis that the nature (or structure) of mathematics achievement is equivalent in children with MIDs and children with RDs during the elementary school years. That is, the findings suggest that the structure of mathematics achievement is unidimensional in both populations. In regard to *Research Question 2*, evidence of configural and metric invariance provides support suggesting that the measurement of mathematics achievement is temporally stable in the combined group model. Inferences concerning change (or growth) in this conceptual domain, however, are inconclusive due to finding intercept non-invariance across the three time points. It was anticipated that intercept non-invariance would emerge in the combined model as a result of student participation in one of two intervention conditions. Specifically, it was thought that students in the mathematics intervention would show an advantage over students who participated in a reading intervention in regard to mean mathematics achievement on the latent scale or with respect to one or more of the indicators. In order to identify sources of intercept non-invariance (and potential sources of intervention effects) follow-up analyses were performed in a 'joint group model'.

Question 1: Follow-up. In the previous analyses, children from both of the larger projects were combined such that children with MIDs and children with RDs who participated in either a reading or mathematics intervention condition were represented in a single group. Subsequently, ME/I was evaluated in this combined group. The results of which suggested that the latent mathematics achievement factor was congeneric and characterized by equivalent factor loadings across time; however, the intercept parameters varied across time. Due to the potential for multiple sources of non-invariance (e.g., between and within group across time), the combined group was separated into two groups for further ME/I evaluation. In this joint group context, groups were differentiated according to intervention condition assignment (reading or mathematics). Therefore, each group consisted of children from both special populations. As with the previous series of analyses, ME/I evaluation in this joint group context involved three primary analyses (i.e., equal form, factor loadings, and intercepts).

The first analysis in the joint group context specified a congeneric model between the two intervention groups across each of the three time points (see Figure 4). Thus, mathematics achievement was specified as one latent construct at each time point for children in the reading intervention group (n = 253) and children in the mathematics intervention group (n = 149). Although correlated residuals and correlations between the latent mathematics achievement factors were specified, for ease in interpretation they are omitted below. Model identification was achieved by fixing the latent mathematics achievement means and variances at each of the three time points to zero in both groups of children. The overall fit for the equal form solution is presented in Table 5. As with previous analyses, the χ^2_{SB} was significant; however, the corresponding fit statistics (see Table 5) provided evidence indicating that a unidimensional measurement model provided a good fit to the data. Thus, evaluation of equal form in the joint group context provided evidence of configural invariance between intervention groups across time. Further, because each intervention group is comprised of children with MIDs and children with RDs, this finding provides additional evidence suggesting that the structure (or nature) of mathematics achievement is fundamentally the same (i.e., unidimensional) in children with MIDs and children with RDs.

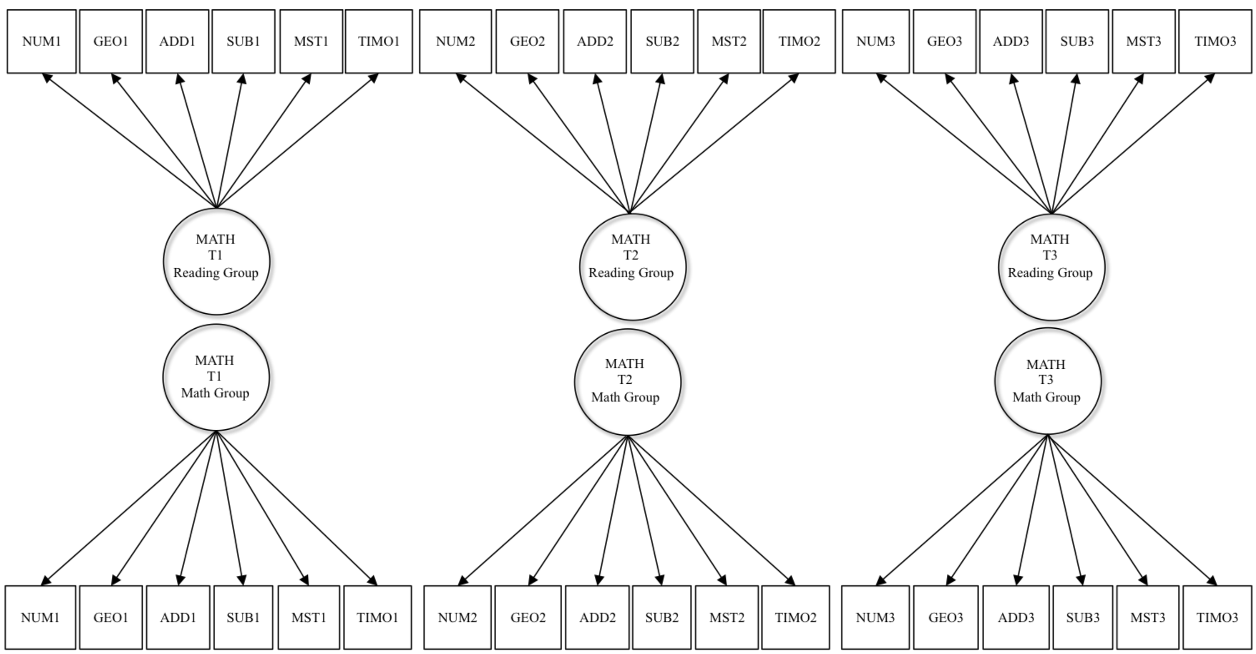

Figure 4. Joint Group Model: Equal Form. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. T1 = Pre-intervention time point, T2 = Mid-intervention time point, T3 = Post-intervention time point. MATH = Mathematics achievement. Read Group = Reading intervention group, Math Group = Mathematics intervention group.

Given evidence of configural invariance, equality of factor loadings between groups for the three time points was investigated in the joint group context. Model identification was achieved by fixing the latent means to zero; however, the metric of mathematics achievement latent variance was set to one for the reading intervention group for the first time point, while freely estimated at the second and third time points. Further, the reading intervention group served as the comparison group. Therefore, the latent variance for mathematics achievement was freely estimated at each of the three time points for the mathematics intervention group. The overall model fit and fit indices for the equal factor loadings solution are presented in Table 5. As displayed, the model fit indices suggest that the equal intercepts solution fits the data well. However, the model χ^2_{SB} of the equal factor loadings solution was 405.55 (*df* = 253), which resulted in a significant difference test, χ^2_{SBdiff} (25) = 46.11, $p < 0.01$.

Table 5. Measurement Invariance Evaluation of Mathematics Achievement: Joint Group Model

	χ^2_{SB} (*df*)	χ^2_{SBdiff}	Δdf	CFI	RMSEA (90% CI)	$P_{RMSEA} \leq .05$	SRMR
Step 1. Equal Form	358.21 (228)***			.981	.053 (.043, .064)	.289	.027
Step 2. Equal Factor Loadings	405.55 (253)***	46.11***	25	.977	.055 (.045, .065)	.204	.037
Step 3. Equal Factor loadings (TIMO loading free in Math Group)	379.19 (250)***	20.10	22	.981	.051 (.040, .061)	.437	.033
Step 4. Equal Intercepts	442.98 (272)***	63.79***	19	.975	.056 (.046, .065)	.145	.039
Step 5. Equal Intercepts: Numeration	379.19 (250)***	00.00	0	.981	.051 (.040, .061)	.437	.033
Step 6. Equal Intercepts: Geometry	389.27 (255)***	10.37	5	.980	.051 (.041, .061)	.406	.033
Step 7. Equal Intercepts: Addition	397.46 (257)***	8.43*	2	.979	.052 (.042, .062)	.346	.034
Step 8. (final model) Equal Intercepts: Subtraction	394.45 (260)***	4.88	5	.980	.051 (.040, .061)	.435	.034
Step 9. Equal Intercepts: Measurement	423.24 (265)***	25.91***	5	.977	.055 (.045, .064)	.212	.036

Note. TIMO = Time/Money; χ^2_{SB} = Satorra-Bentler scaled χ^2; χ^2_{SBdiff} = difference test; Δdf = change in degrees of freedom; CFI = Comparative fit index, RMSEA = Root mean square error of approximation, $P_{RMSEA} \leq .05$ = Test of close fit, SRMR = Standardized root mean square residual.
$^{*}p < .05$; $^{***}p < .001$.

Preliminary analyses of within group invariance (see Appendix 3) for the mathematics intervention group identified the Time/Money factor loadings as evidencing temporal non-invariance. The subsequent model (Step 3) therefore released the corresponding constraint and allowed the Time/Money factor loading to vary across time within the mathematics intervention group, but not the reading intervention group. Model identification was achieved as in the previous analysis. This model specification provided a good fit to the data as characterized by the model fit indices. Moreover, compared to the equal form model, the addition of the factor loadings constraint (with Time/Money factor loading "free" for the mathematics group) did not significantly reduce model fit, χ^2_{SBdiff} (22) = 20.10, p = 0.53. Thus, five of the six factor loadings were fully invariant between groups across time.

As displayed in Figure 5 below, the factor loadings are the same for the reading and mathematics intervention groups at each time point for the Numeration (2.84), Geometry (2.84), Addition (2.39), Subtraction (1.74), and Measurement (2.68) subtests. The factor loadings for the Time/Money subtest, however, are the same at each of the three time points for the reading intervention group (2.11), while changing at each time point for the mathematics intervention group (1.92, 2.04, 2.66, for pre-, mid-, and post-intervention, respectively). Non-invariance of the Time/Money factor loading is further specified in Figure 5 by the red intercept parameter with asterisk. Non-invariance of the Time/Money subtest indicates that with each unit change in the latent mathematics achievement construct, the Time/Money subtest is expected to change differentially by time point for the mathematics intervention group, but not the reading intervention group. However, all other factor loadings are stable between groups across time. Thus, with one unit change in the latent mathematics achievement construct, the Numeration subtest is expected to change 2.84 units, the Geometry subtest, 2.84 units, the Addition subtest, 2.39 units, and so on. Moreover, expected change in the indicators is consistent across groups. Given this partially invariant factor loadings model, Figure 5 also includes estimates of the variance for the latent mathematics achievement construct for both intervention groups at each time point. Finally, note that as a result of non-invariance of the Time/Money factor loading within the mathematics intervention group, between group comparisons of Time/Money intercepts cannot be investigated.

Given partial invariance of factor loadings in the joint context, tests of equal intercepts were investigated. Model identification was achieved by fixing the latent variance and means to one and zero, respectively, for the reading intervention group. The remaining latent variances and means for the reading intervention group were freely estimated while all three latent variances and means for the mathematics intervention group were freely estimated. The model results and difference testing (see Table 5) indicated that the addition of the intercepts' equality constraint between groups across time for the Numeration, Geometry, Addition, Subtraction, and Measurement subtests (intercept equivalence was not evaluated for Time/Money due to factor loading non-invariance in the previous analysis) significantly reduced model fit. The model χ^2_{SB} of the equal intercepts solution was 442.98(df = 272), which resulted in a significant χ^2 difference test, χ^2_{SBdiff} (19) = 63.79, $p < 0.001$, suggesting that the scale of the latent mathematics achievement is unstable (or inconsistent) between groups across time. In attempt to identify the specific source(s) of intercept non-invariance, subsequent analyses systematically examined intercept equivalence between groups across time one indicator at-a-time in a step-wise fashion.

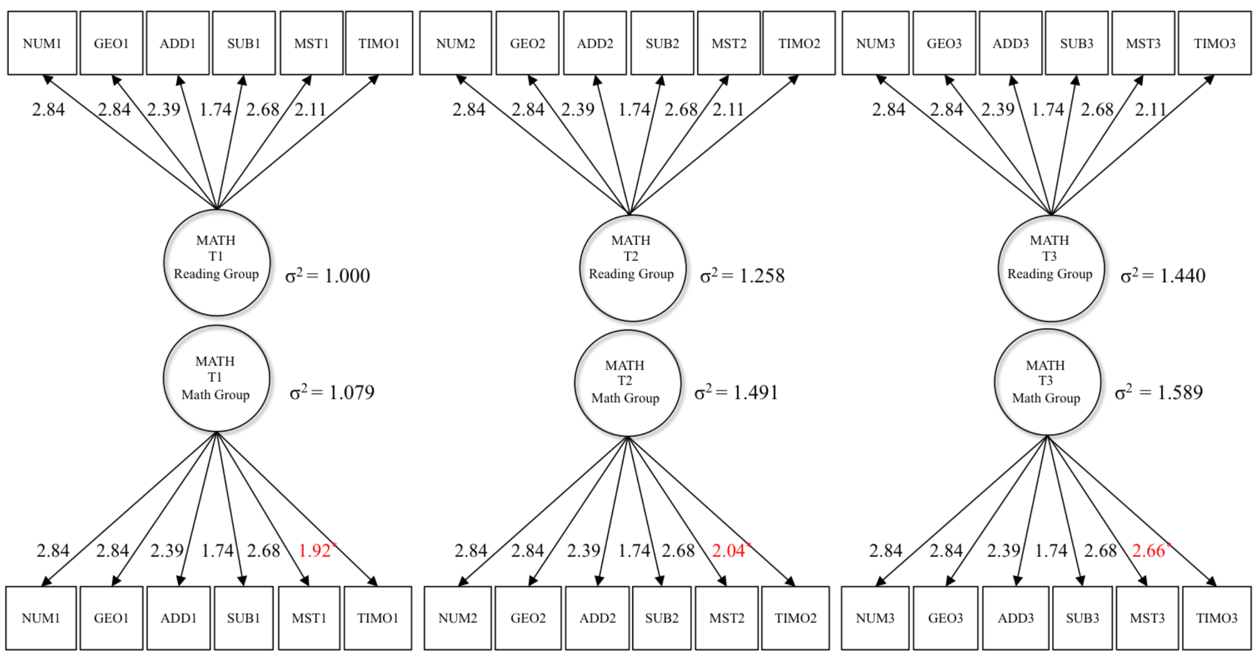

Figure 5. Joint Group Model: Equivalent Factor Loadings. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point. σ^2 = factor variance. MATH = Mathematics achievement. Read Group = Reading intervention group, Math Group = Mathematics intervention group. Factor loadings displayed *inside* of the figure. Asterisk indicates non-invariance.

As displayed in Table 5, these analyses built from the equal factor loadings model where the Time/Money factor loading was allowed to vary across time within the mathematics intervention group (Step 3). In Steps 5 through 9, the hypothesis that each indicator's (Numeration, Geometry, Addition, Subtraction, Measurement) intercept was invariant between groups and across time was investigated. Of the five intercepts, the results of ME/I analyses suggested that the Numeration, (Step 5), Geometry (Step 6), and Subtraction (Step 8) intercepts were invariant between groups across time. Thus, the means for Numeration, 6.16, Geometry, 5.39, and Subtraction, 1.97, were consistent between groups across time. In contrast, ME/I analyses suggested that the Addition (Step 7) and Measurement (Step 9) intercepts were non-invariant between groups across time as evidenced by significantly reduced model fit compared to the previously specified model. Thus, the means for the Addition and Measurement subtests were inconsistent between groups across time. The final model is displayed in Figure 6. As in previous figures, residual and latent factor correlations are omitted for ease in interpretation (see Appendix 4 for final model residual and latent correlations). Figure 6 also includes estimates for the latent mathematics achievement means between groups for each time points. In general, intercept non-invariance suggests that one group evidences an advantage on a given subtest; however, in examining Figure 6, between group differences appear to be minimal. Rather than a group advantage, intercept non-invariance likely is due to within group measurement inconsistency. For instance, in examining Figure 7, the Addition intercepts appear to be consistent within the mathematics intervention group and between groups at the pre-intervention time point. However, the Addition intercepts evidence a marked increase between the pre- and mid-intervention time points within the reading intervention group. For the Measurement subtest, intercepts for the reading group appear stable across time and between groups at time three and perhaps, time two. Thus, intercept non-invariance for the Measurement subtest is likely due to the marked decrease in this parameter estimate from time one (4.60) to time two (3.78) within the mathematics group. In summary, ME/I evaluation of mathematics achievement demonstrated that the latent construct was characterized by configural invariance (i.e., equal form) and partial metric (i.e., factor loadings) and scalar (i.e., intercept) invariance between the two intervention groups across time. The final model parameter estimates for each indicator’s factor loading, intercept, and residual variance by intervention group and time point are displayed in Table 6.

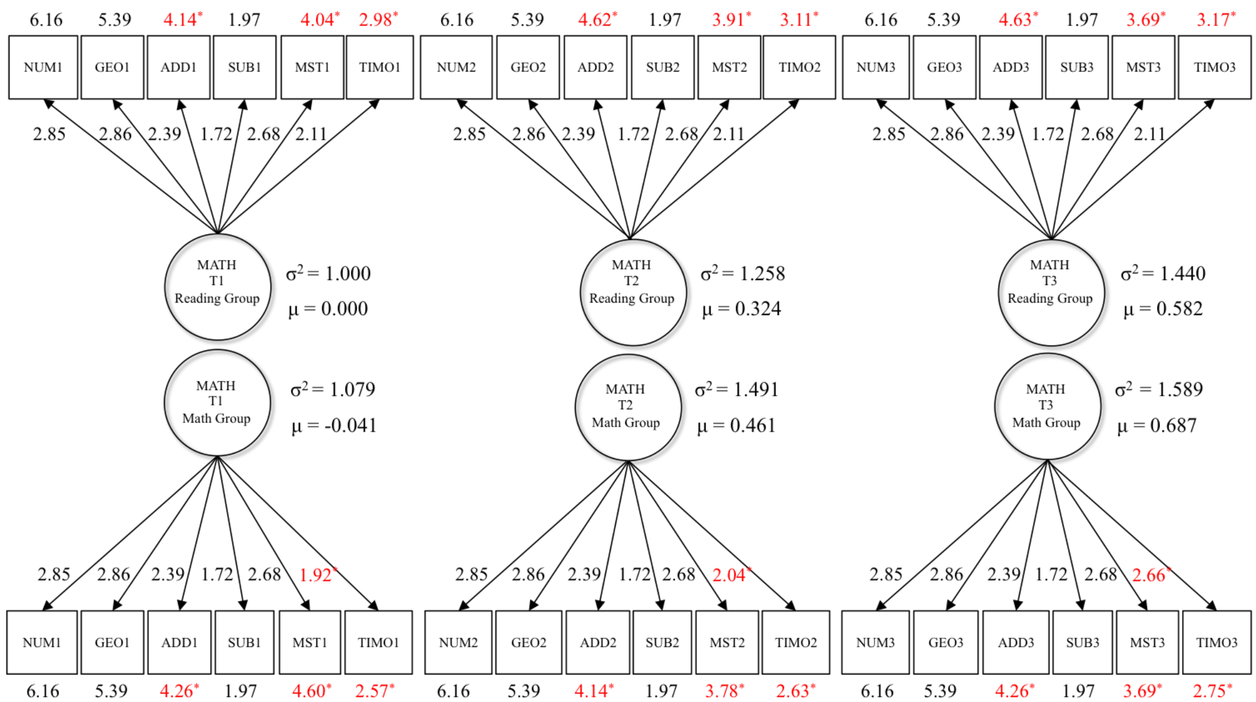

Figure 6. Joint Group Model: Equivalent Intercepts. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point. σ^2 = factor variances, μ = factor means. MATH = Mathematics achievement. Read Group = Reading intervention group, Math Group = Mathematics intervention group. Factor loadings displayed *inside* of the figure. Intercepts displayed *outside* each indicator. Asterisk indicates non-invariance.

To better understand the non-invariance that characterized the Addition, Measurement, and Time/Money intercepts, longitudinal plots of the intercepts differentiated by intervention condition are displayed below. In regard to the Addition (Figure 7) and Measurement (Figure 8) subtests, they were characterized by equivalent factor loadings between groups across the three time points. As displayed in Figure 7, the reading and mathematics group evidenced similar intercepts for the Addition subtest at the pre-intervention time point; however, the means diverge at the intervention mid-point (difference of 0.48) and maintain an attenuated difference at post-intervention (difference of 0.37). In contrast, Figure 8 indicates that there was a pre-existing group difference in mean achievement for the Measurement indicator that provided an initial advantage for the mathematics group over the reading group (difference of 0.56). By the second and third time point, however, this advantage was negligible (difference of 0.13 and 0.001, for the second and third time points, respectively). With respect to the final indicator, Time/Money (see Figure 9), this was the only subtest that evidenced temporal non-invariance for the factor loadings (i.e., for the mathematics group). Consequently, the intercept plots are provided for descriptive purposes only as the metric of measurement varies between groups.

In summary, configural, metric, and scalar invariance were evaluated in a joint group context that differentiated children with MIDs and children with RDs according to their intervention condition participation. Sources of indicator non-invariance were identified. It was demonstrated that a congeneric factor model (i.e., equal form) provided the best representation of the underlying factor structure between groups across time. Evidence of metric invariance (i.e., factor loading equivalence) also was established suggesting that the indicators evidenced comparable relationships to the latent mathematics achievement construct between groups across time (with the exception of Time/Money loading in the mathematics group). Evidence of partial scalar invariance also was obtained. Specifically, the intercepts for Numeration, Geometry, and Subtraction were temporally stable between groups across time, while the intercepts for Addition, Measurement, and Time/Money were characterized by non-invariance. In responding to *Research Question 1*, the results provided consistent evidence of configural invariance indicating that the latent mathematics achievement construct was characterized by one latent factor. The absence of configural invariance in the combined or joint group context at the pre-intervention time point would suggest that the structure of mathematics achievement in children with MIDs was fundamentally different than that in children with RDs. This difference likely would have manifested as poor model fit and unacceptable model fit statistics for the configural invariance analyses. As displayed in Tables 4 and 5, model fit indices concerned with equal form provided support for configural invariance. Thus, the hypothesis that the nature (or structure) of mathematics achievement is equivalent in children with MIDs and children with RDs is supported.

Table 6. Joint Group Model: Final Model Parameter Estimates

		Factor Loadings (SD)		*Factor Intercepts (SD)*		*Residual Variances (SD)*	
Group		Reading	Math	Reading	Math	Reading	Math
Math1	Numeration	2.85 (.13)	2.85 (.13)	6.16 (.19)	6.16 (.19)	2.03 (.28)	1.80 (.30)
	Geometry	2.86 (.13)	2.86 (.13)	5.39 (.21)	5.39 (.21)	6.44 (.59)	6.22 (.82)
	Addition	2.39 (.12)	2.39 (.12)	**4.14 (.20)**	**4.26 (.21)**	3.22 (.45)	1.68 (.26)
	Subtraction	1.72 (.09)	1.72 (.09)	1.97 (.12)	1.97 (.12)	1.38 (.18)	0.85 (.12)
	Measurement	2.69 (.13)	2.69 (.13)	**4.04 (.21)**	**4.60 (.24)**	3.26 (.41)	3.06 (.41)
	Time/Money	2.11 (.15)†	**1.92 (.14)**	**2.98 (.16)**	**2.57 (.17)**	2.39 (.25)	1.76 (.26)
Math2	Numeration	2.85 (.13)	2.85 (.13)	6.16 (.19)	6.16 (.19)	2.22 (.27)	1.37 (.22)
	Geometry	2.86 (.13)	2.86 (.13)	5.39 (.21)	5.39 (.21)	6.83 (.64)	7.01 (.88)
	Addition	2.39 (.12)	2.39 (.12)	**4.62 (.21)**	**4.14 (.22)**	3.59 (.41)	2.40 (.32)
	Subtraction	1.72 (.09)	1.72 (.09)	1.97 (.12)	1.97 (.12)	2.01 (.24)	1.47 (.26)
	Measurement	2.69 (.13)	2.69 (.13)	**3.91 (.21)**	**3.78 (.25)**	3.27 (.38)	3.79 (.51)
	Time/Money	2.11 (.15)†	**2.04 (.16)**	**3.11 (.18)**	**2.63 (.18)**	2.64 (.34)	2.02 (.35)
Math3	Numeration	2.85 (.13)	2.85 (.13)	6.16 (.19)	6.16 (.19)	2.49 (.54)	2.59 (.40)
	Geometry	2.86 (.13)	2.86 (.13)	5.39 (.21)	5.39 (.21)	5.81 (.60)	4.85 (.58)
	Addition	2.39 (.12)	2.39 (.12)	**4.63 (.22)**	**4.26 (.22)**	4.12 (.48)	3.05 (.45)
	Subtraction	1.72 (.09)	1.72 (.09)	1.97 (.12)	1.97 (.12)	1.76 (.21)	2.04 (.43)
	Measurement	2.69 (.13)	2.69 (.13)	**3.69 (.22)**	**3.69 (.26)**	3.35 (.45)	3.97 (.52)
	Time/Money	2.11 (.15)†	**2.66 (.22)**	**3.17 (.19)**	**2.75 (.23)**	4.06 (.49)	3.16 (.62)

Note. Math1 = Pre-intervention, Math2 = Mid-point of intervention, Math3 = Post-intervention; Model fit statistics are χ^2_{SB} (260) = 394.45, p = .001, CFI = .98, RMSEA = .051 (90% CI = .040-.061), SRMR = .034; † = Constrained factor loading within the reading group; bold indicates a freely varying parameter.

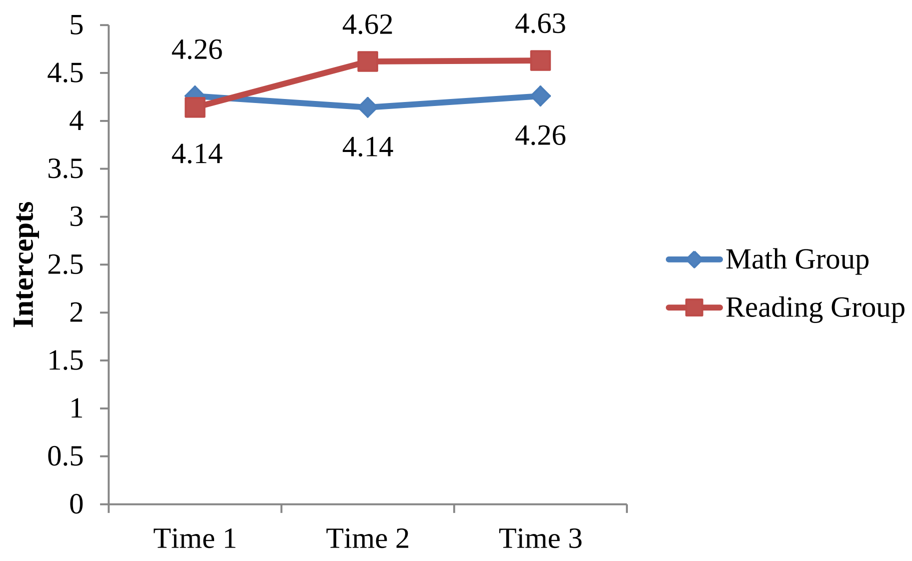

Figure 7. Addition Intercepts by Intervention Group. Time 1 = Pre-intervention, Time 2 = Intervention mid-point, Time 3 = Post-intervention.

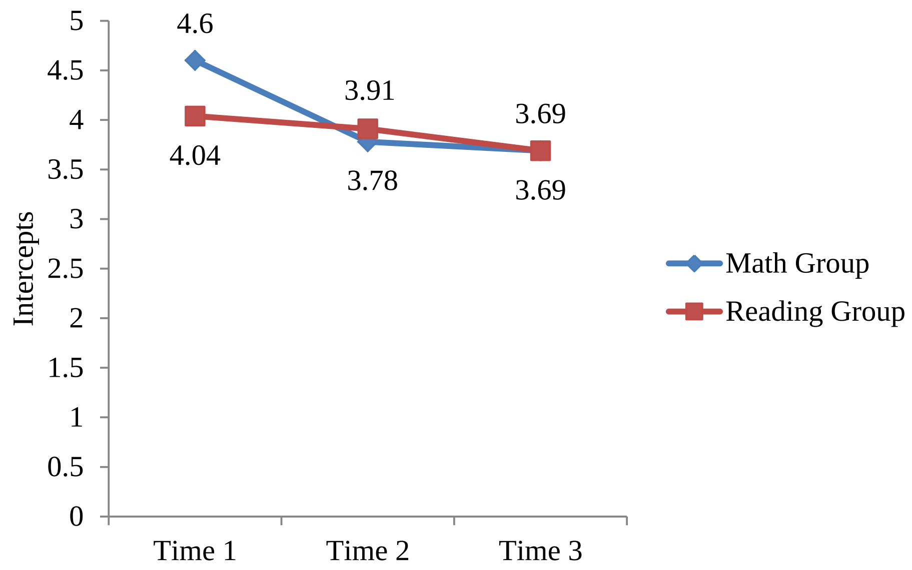

Figure 8. Measurement Intercepts by Intervention Group. Time 1 = Pre-intervention, Time 2 = Intervention mid-point, Time 3 = Post-intervention.

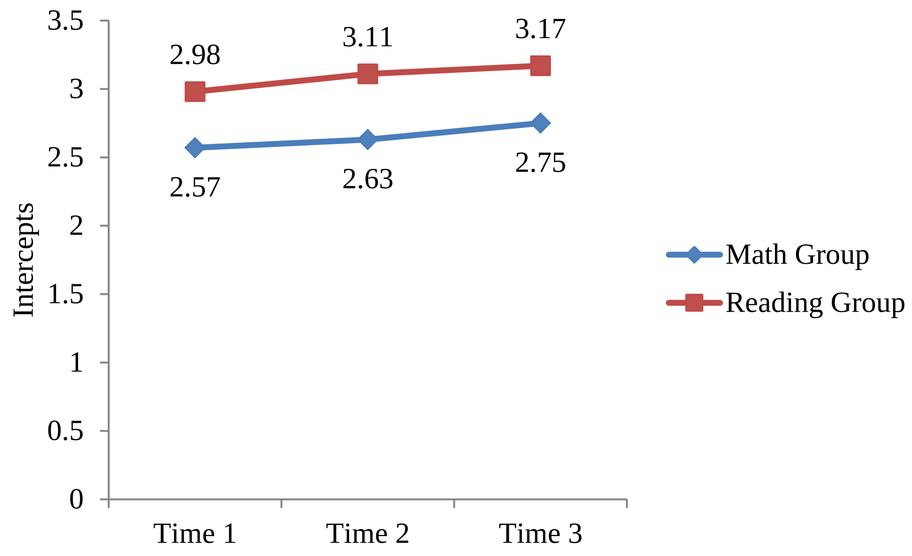

Figure 9. Time/Money Intercepts by Intervention Group. Time 1 = Pre-intervention, Time 2 = Intervention mid-point, Time 3 = Post-intervention.

In responding to *Research Question 2*, the results provided consistent evidence of metric invariance. In the combined group model, the results indicated that the factor loadings for all six mathematics indicators were monotonically invariant (i.e., consistently increasing) and proportional. In contrast, the results of the joint group model indicated that the factor loadings for the six indicators were monotonically invariant and proportional in both intervention groups, with the exception of the Time/Money factor loadings in the mathematics group, which were characterized by non-invariance. This source of non-invariance is likely due to the substantial increase in the Time/Money factor loading from mid- to post-intervention. As result, the amount of predicted change on the Time/Money subtest given a unit change in the latent mathematics achievement factor varied across time within the mathematics group. In regard to intercept invariance in the combined group model, the results indicated that the addition of the intercept constraint significantly reduced model fit (see Table 4). Therefore, intercept invariance was evaluated within the context of the joint group model where students were differentiated according to their participation in a mathematics or a reading intervention condition. The results of this series of analyses suggested that three intercepts (Numeration, Geometry, and Subtraction) were temporally stable between groups across time, while three others (Addition, Measurement, Time/Money) were not. Thus, the collective results indicate that the mathematics achievement factor is characterized by the same configuration and pattern of indicator-factor loadings. Moreover, half of the intercepts demonstrated longitudinal stability between groups, while the non-invariance evidenced does not appear to be due to intervention effects (see Figures 7 and 8); as the indicator means do not favor students in the mathematics intervention condition. Thus, the hypothesis that the measurement of mathematics achievement is equivalent in children with MIDs and children with RDs, who were differentiated according to intervention condition participation, is supported as evidenced by the strong, partially invariant measurement model.

In an effort to further substantiate the previous results and rule out an alternative hypothesis, an additional analysis was run. For this analysis, a fully invariant measurement model (omnibus test) that differentiated children according to disability status *and* intervention assignment was specified. This specification resulted in four separate groups. For children with MID, one group consisted of 83 children that participated in the mathematics intervention, while the other consisted of 182 that participated in a reading intervention. For children with RD, one group consisted of 65 children that participated in the mathematics intervention, while the other consisted of 70 children that participated in a reading intervention. Evaluation of model fit and approximate fit indices, χ^2_{SB} (*df*) = 1023.26 (566), $p < .001$, CFI = 0.916, RMSEA (90% CI) = 0.090 (0.81, 0.099), $p_{RMSEA} \leq .05 < .0001$, SRMR = 0.111, did not support the tenability of this model over those reported in detail in this study. Therefore, the four group model specification was rejected as a tenable solution in favor of the two group approach with children differentiated by intervention condition participation.

In conclusion, the results of ME/I evaluation provide evidence of a strong, partially invariant model of mathematics achievement for children with MIDs and children with RDs differentiated according to intervention condition. Specifically, strong factorial invariance held for a subset of the measured indicators (Numeration, Geometry, Subtraction), whereas partial invariance held for the other subset (Addition, Measurement, Time/Money). It therefore follows that a unit change in the latent mathematics achievement construct is associated with comparable changes between groups across time.

Further, the collective results demonstrate that mean change over time in the latent mathematics achievement construct is due to true change and not changes in the structure or measurement of mathematics achievement (for discussion on true change, see Golembiewski, Billingsley, & Yeager, 1976; Vandenberg & Lance, 2000). Consequently, inferences and conclusions concerning group differences and mathematics achievement growth within the present study are meaningful.

3.6 Evaluating Students' Response to Mathematics Intervention

Question 3. Evaluation of the structure (or nature) of mathematics achievement and its measurement provided evidence for a strong, partially invariant model of mathematics achievement. Because the measurement of mathematics achievement was reliable and valid, inferences and conclusions concerning growth and subsequently, evaluation of students' response to a mathematics intervention are meaningful. Therefore, *Research Question 3*, which examines mathematics achievement growth, was investigated. As discussed, students with MIDs and students with RDs comprised each intervention condition. Further, whereas the previous analyses were concerned with the individual indicators (KeyMath-R subtests), analyses investigating response to intervention are focused on mathematics achievement as measured on the latent scale. Latent mathematics achievement includes information from all six indicators because configural and partial metric and scalar invariance was established.

In order to investigate mathematics achievement growth and make group comparisons, the reading intervention group was specified as the reference group by fixing their latent mean at the pre-intervention time point to 0.0 (see Figure 10). All other latent means were freely estimated. Comparison of the groups, then, was based on the difference from zero on the latent scale. It was originally projected that latent growth curves would be estimated separately for each intervention group. However, preliminary analyses suggested that mathematics achievement growth was curvilinear for the mathematics intervention group. Linear growth curves for the latent mathematics achievement construct were therefore not specified in order to avoid model misfit as evidenced by model fit indices test statistics.

Although latent growth curves were not estimated, group differences for the latent means were tested. Difference testing of latent mathematics achievement means between groups at the second and third time points, was carried out using M*plus* software. The results suggest that children in the mathematics intervention group evidenced an advantage over those in the reading intervention group at mid-intervention, difference = 0.137 (0.13), p = 0.29, and post-intervention, difference = 0.105 (0.14), p = 0.44. Although, these mathematics achievement mean differences were not statistically significant, they are in latent score units and therefore, can be interpreted as standardized deviations relative to mathematics achievement at the pre-intervention time point. Therefore, the mathematics group's latent mean (μ = -0.041) represents a deviation from the reference group and indicates that, on average, students in the mathematics intervention group scored 0.041 units *lower* than their peers in the reading intervention group. Latent change scores were then calculated by taking the difference of each group's post- and pre-intervention latent means (0.582 – 0.00 = 0.582, and 0.687 – [-0.041] = 0.728, for the reading and mathematics group, respectively) the results indicate that, in addition to evidencing a small advantage at mid- and post-intervention assessment, the mathematics group demonstrated greater growth (0.728) compared to the reading group (0.582) over the course of the intervention (difference = 0.146). Thus, on average, students that participated in the mathematics intervention condition started out with lower mathematics achievement scores compared to those in the reading intervention conditions. However, on average, by the end of the school year (and completion of the students respective interventions), students in the mathematics group outperformed those in the reading group. In short, their performance caught up with and surpassed that of their peers in the reading intervention group.

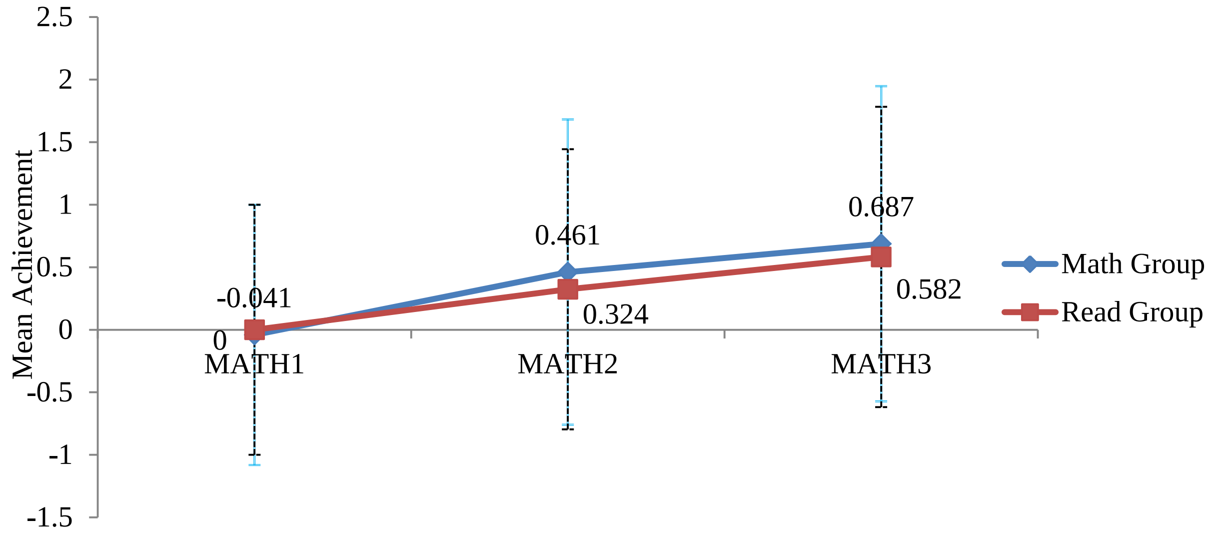

Figure 10. Longitudinal Mathematics Achievement by Intervention Group. Scores reported are latent means (i.e., kappa) with standard error bars; MATH = Mathematics achievement; 1 = Pre-intervention, 2 = Intervention mid-point, 3 = Post-intervention. Group = Intervention group.

3.7 Treatment Effects

In considering treatment effects (or effect size) in the present study, the reported growth parameters were 0.582 and 0.728 for the reading and mathematics groups, respectively. These latent growth parameters indicate that, on average, students in each intervention condition evidenced improved mathematics achievement over the school year with the students in the mathematics group showing an advantage, 0.146, over students in the reading group. The effect size reported in the present study is somewhat stronger than that reported by McKenzie, Marchand-Martella, Martella, and Moore (2005) who utilized CMC Level K to instruct preschool children. Their 16 participants included children with (n = 5) and without (n = 11) developmental delays; each of which completed all 30 lessons of CMC Level K (Engelmann & Becker, 1995). The results of McKenzie et al. (2005) indicated that the group of students with developmental delay evidenced an effect size of 0.54, whereas the group of students without developmental delays evidenced an effect size of 0.61, on the Battelle Developmental Inventory (Newborg et al., 1984). In addition to the effect sizes reported in the present study being somewhat stronger than that reported in McKenzie et al. (2005), it also is important to note that the presently reported estimates are corrected for measurement error and are therefore, unbiased estimates. That is, the effect sizes reported in McKenzie et al. (2005) were not corrected for measurement error and are therefore biased estimates. Consequently, they are not necessarily reliable estimates of treatment effects in the population from which the sample was drawn.

In an attempt to compare the present effect size estimates with that of the discussed literature (i.e., Van Luit & Naglieri, 1999; Van Luit & Schopman, 2000), Cohen's *d* with pooled variance (represented in the denominator) was estimated for *overall raw mathematics achievement*. This variable was created by taking the sum of each participant's scores for the six KeyMath-R subtests at the pre- and post-intervention time points, respectively. Table 7 displays the data used to estimate Cohen's *d* and the procedures for calculating effect size as found in Van Luit and Naglieri (1999) and Van Luit and Schopman (2000). In comparison to the cited literature, the magnitudes of the effect sizes reported in the present study are weaker. This finding is likely due in part to the use of a norm-referenced measure of mathematics achievement as opposed to a researcher designed test. Norm-referenced assessments can lack sensitivity to detect subtle changes in an academic domain, especially when the target measure (i.e., the KeyMath-R) fails to include children with disabilities in the standardization procedures (see the KeyMath-R examiner's manual). Thus, it is not uncommon for treatment effect size estimates for criterion-referenced measures (teacher or researcher developed) to be stronger in magnitude compared to effect size estimates for norm-referenced measures (Berkeley, Scruggs, & Mastropieri, 2010). Further explanation concerning the effect size differences between the present study and the cited literature is presented in the Discussion section below.

Table 7. Pre- and Post-Intervention Means, Standard Deviations, and Effect Sizes Across Studies

	Pre-Intervention			Post-Intervention			
Study	*n*	*M*	*SD*	*n*	*M*	*SD*	Pre-Post Effect[a]
Van Luit & Naglieri (1999)							
Intervention	42	11.3	6.5	42	31.9	5.4	3.45
Comparison	42	12.5	6.7	42	18.2	8.2	0.76
Van Luit & Schopman (2000)							
Intervention	62	46.1	9.3	62	59.5	9.3	1.44
Comparison	62	46.9	9.3	62	53.3	9.4	0.68
Present Study: Foster (2014)							
Intervention	149	24.50	15.53	143	35.23	19.98	0.60
Comparison	246	24.25	15.26	238	34.32	18.05	0.60

Note. [a]Effect size = $\frac{M_2 - M_1}{\sqrt{\frac{n_2 \times SD_2^2 + n_1 \times SD_1^1}{n_2 + n_1}}}$

3.8 Differential Indicator Bias

In an attempt to describe the potential differential effects of the non-invariant indicators, the expected observed scores for the mathematics and reading intervention groups were evaluated. In the present context, differential item functioning refers to the between group difference in observed scores for a given mathematics subtest when the groups have the same value of the underlying attribute (McDonald, 1999). Thus, students in the mathematics and reading groups who have the same common factor score can be expected to have different observed scores on the non-invariant indicators (i.e., Addition, Measurement, Time/Money). The following is the indicator-specific equation used to examine differential subtest functioning, $Y_{gmt} = \tau + \lambda(\eta) + \varepsilon$, where *Y* is the expected observed score for an individual in intervention group $_{g\text{(Math = 0; Reading = 1)}}$, on subtest $_{m\text{(A = Addition; M = Measurement, T = Time/Money)}}$, for time point $_{t\text{(3 = post-intervention)}}$. Tau, τ, represents the group, measure, and time specific intercept; lambda, λ, the group, measure and time specific factor loading; eta, η, the strength of the underlying attribute; and epsilon, ε, the residual effect that is assumed to be zero. Thus, using the intercept and factor loadings for the non-invariant indicators at the post-intervention time point, while setting eta (η) to -1, 0, +1, respectively, results in the equations and predicted observed scores displayed in Table 8.

As displayed in Table 8, of the three non-invariant indicators, the observed scores for the Measurement subtest at the post-intervention time point appears to be comparable between groups (Y_{1M3} and Y_{0M3} = 1.00, 3.69, 6.38 when η = -1, 0, +1, respectively). Therefore, non-invariance of this indicator is likely due to the between group difference at the first time point, or the within group difference for the mathematics group from the first to second time points (see Table 6); the latter of which appears to be the largest in magnitude. With respect to the predicted observed scores for the Addition subtest at post-intervention (see Figure 11), the results suggest that the reading group evidenced a slight advantage at each value (-1, 0, +1) of the underlying latent factor (i.e., η). Finally, for Time/Money, there appears to be an interaction such that when η = -1 and when η = 0, there is an advantage for the reading group, while at eta +1, there is a slight advantage for the mathematics group; however, this latter finding must be interpreted with caution as the Time/Money factor loading was freely estimated in the mathematics group but not the reading group. Consequently, the metric for the Time/Money subtest, although relative to the latent scale, is not necessarily equivalent across the intervention groups. In summary, differential functioning of the non-invariant subtests is minimal and likely does not interfere with using the KeyMath-R to measure mathematics achievement growth. Moreover, the displayed discrepancies are likely not of sufficient magnitude to interfere with the use of the KeyMath-R in the groups being compared (see Millsap, 2005).

Table 8. Predicted Observed Scores by Group for Non-Invariant Indicators at Post-Intervention

	Reading Group	Mathematics Group
Addition	Y_{1A3} = 4.63 + 2.39(-1) + 0 = 2.24	Y_{0A3} = 4.26 + 2.39(-1) + 0 = 1.87
	Y_{1A3} = 4.63 + 2.39(0) + 0 = 4.63	Y_{0A3} = 4.26 + 2.39(0) + 0 = 4.26
	Y_{1A3} = 4.63 + 2.39(+1) + 0 = 7.02	Y_{0A3} = 4.26 + 2.39(+1) + 0 = 6.65
	Y_{1M3} = 3.69 + 2.69(-1) + 0 = 1.00	Y_{0M3} = 3.69 + 2.69(-1) + 0 = 1.00
Measurement	Y_{1M3} = 3.69 + 2.69(0) + 0 = 3.69	Y_{0M3} = 3.69 + 2.69(0) + 0 = 3.69
	Y_{1M3} = 3.69 + 2.69(+1) + 0 = 6.38	Y_{0M3} = 3.69 + 2.69(+1) + 0 = 6.38
	Y_{1T3} = 3.17 + 2.11(-1) + 0 = 1.06	Y_{0T3} = 2.75 + 2.66(-1) + 0 = 0.09
Time/Money	Y_{1T3} = 3.17 + 2.11(0) + 0 = 3.17	Y_{0T3} = 2.75 + 2.66(0) + 0 = 2.75
	Y_{1T3} = 3.17 + 2.11(+1) + 0 = 5.28	Y_{0T3} = 2.75 + 2.66(+1) + 0 = 5.41

Note. Indicator-specific equations, $Y_{gmt} = \tau + \lambda(\eta) + \varepsilon$, *Y* is the expected observed score for an individual in the math (*0*) or reading (*1*) group, on the Addition (*A*), Measurement (*M*), or Time/Money (*T*) subtest, for the post-intervention time point (*3*). Tau (τ) represents the group, measure, and time specific intercept; lambda (λ) represents the group, measure, and time specific factor loading; eta (η) represents the factor loading (or underlying attribute); and epsilon (ε) represents residual effects and is assumed to be zero.

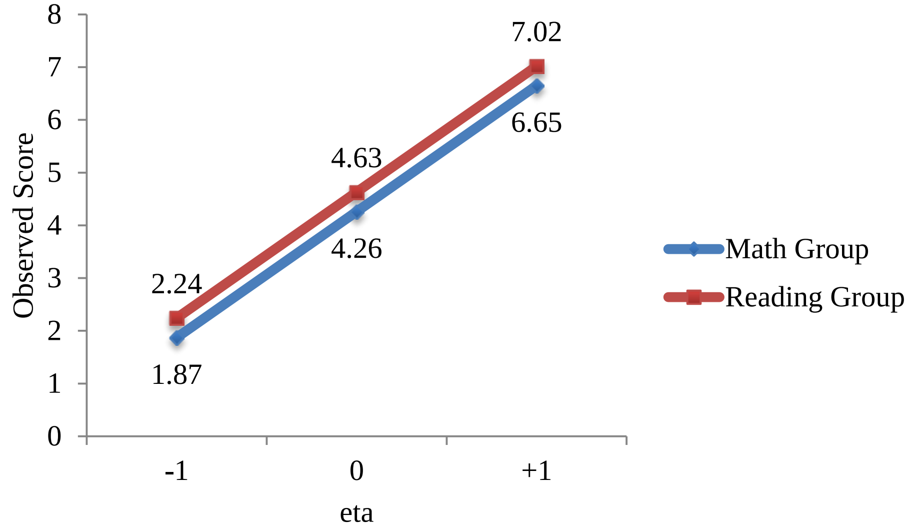

Figure 11. Predicted Post-Intervention Observed Scores for Addition by Intervention Group. Eta (η) represents the value of the underlying attribute.

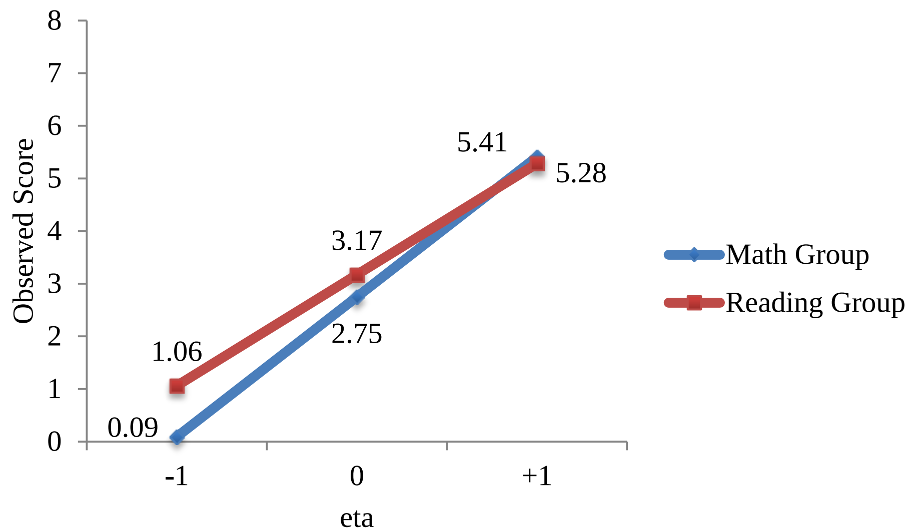

Figure 12. Predicted Post-Intervention Observed Scores for Time/Money by Intervention Group. Eta (η) represents the value of the underlying attribute.

In summary, difference testing of latent means, although not significant, provided evidence that indicates that students who participated in the mathematics intervention condition showed stronger mathematics achievement at mid- and post-intervention compared to their peers who participated in a reading intervention condition. Further, comparisons of mathematics achievement growth from pre- to post-intervention provided evidence indicating that, on average, students in the mathematics intervention group improved more than their peers in the reading intervention group. In regard to effect size, the present estimates are corrected for measurement error and provide evidence of a small treatment effect according to the latent scale over time, 0.146 latent units, in favor of the mathematics group. Finally, evaluation of differential subtest functioning suggests that non-invariance evidenced at post-intervention is minimal and likely does not interfere with reliability and validity of using the KeyMath-R to document mathematics achievement growth in children with MIDs and children with RDs. Therefore, in response to *Research Question 3*, the results provide support for the conclusion that students responded favorably to the mathematics intervention, albeit, a small effect.

4. DISCUSSION

The overarching goal of this dissertation was to examine the nature of mathematics achievement and response to intervention in children with MIDs and children with RDs. In doing so, the present study extends the mathematics achievement literature in several ways. To begin with, students with MIDs have only been included in a few relatively large randomized control studies related to mathematics. Consequently, very little is known about their mathematical development (Branakaer, Ghesquière, & De Smedt, 2011; Foster et al., 2014) and their response to intervention. Furthermore, the present study is the first to systematically examine the structure of mathematics achievement in children with MIDs and children with RDs. Additionally, mathematics achievement growth was investigated after establishing a reliable and valid measurement model, which is rarely completed in applied research (Vandenberg & Lance, 2000). Finally, the research methodology employed, CFA with ME/I evaluation, provides advantages (e.g., correction for measurement error) that traditional methods (i.e., ANOVA) cannot (see Vandenberg & Lance, 2000).

4.1 The Structure of Mathematics Achievement

Using six subtests of the KeyMath-R (i.e., Numeration, Geometry, Addition, Subtraction, Measurement, Time/Money), the results of this study confirm the hypothesis that the structure (and therefore, nature) of mathematics achievement is equivalent in elementary aged children with MIDs and children with RDs. Specifically, the structure of mathematics achievement was evaluated in the combined group and joint group context. In the former context, the single combined group consisted of children from both special populations and intervention conditions, whereas in the latter context, groups were differentiated according to intervention condition participation (not disability status). Results for each model supported the hypothesis that the underlying factor structure of mathematics achievement was unidimensional. The unidimensional model provides a parsimonious and substantively meaningful model of mathematics achievement; that early mathematical development is comprised of a set of highly interrelated skills. Further, because the equal form model specification fit the data well in the combined group and the joint group context, it can be concluded that the structure of mathematics achievement is equivalent in children with MIDs and children with RDs. Had the equal form model specification not fit the data well (in either series of models), additional analyses would have been carried out to identify sources of non-invariance. Such analyses may have led to differentiating students according to intervention condition assignment *and* disability status.

These present results are consistent with previous research (Parmar et al., 1994) that *concluded* that students with MR and students with LD are not qualitatively different despite, students with RDs, on average, demonstrating stronger performance than students with MIDs on several measures of mathematics achievement (see Appendix 1). Taken together, the results of the present study and those of Parmar et al. (1994) indicate that disability group differences are only quantitative in nature. Thus, mathematics achievement in these two special populations is fundamentally the same. It is not the case, for example, that the structure of mathematics achievement in children with MIDs is best described as consisting of one domain (e.g., global mathematics achievement), whereas for children with RDs, this conceptual domain is best described as consisting of two areas (e.g., basic concepts and problem solving/reasoning). Moreover, because of the longitudinal design employed in the present study, it can be concluded that children with MIDs are developing mathematics achievement in the same manner, with the same structure, as their peers with RDs; and given that children with RDs evidence mathematics development that is most closely related to typically achieving children (i.e., of children with MD, RD, and MDRD), children with MIDs and RDs likely are developing mathematics achievement with the same structure as their typically achieving peers.

Other measurement characteristics of mathematics achievement, metric and scalar invariance were also evaluated in the present study. With regard to the former, evaluation of the factor loadings established that the metric of measurement in mathematics achievement was largely equivalent. That is, factor-loading parameters (except for Time/Money) were in the same order of magnitude (i.e., monotonically invariant) and proportional between groups for each of the three time points. Non-invariance that characterized the factor loading for the Time/Money subtest in the mathematics group may have been due to floor effects present at the pre-intervention time point. That is, students in the mathematics intervention condition, on average, performed below their peers at each of the three time points and a substantial number of students has a score of 0 at the baseline time point; however, as students in the mathematics condition improved on the Time/Money subtest at mid- and post-intervention (i.e., the number of students with a score of 0 decreased; see Tables 1, 2, and 3), the factor loadings increased in strength and variability over time, despite students showing minimal improvement on this subtest.

With regard to scalar invariance, the lack of intercept non-invariance in the combined group model provided the impetus to evaluate ME/I in the joint group context. Within the latter model, students were differentiated by intervention condition participation. Partial ME/I was then pursued because instructional effects can show up as differences in measurement parameters, especially parameters concerned with mean achievement between groups (McArdle, 1996). The results of partial ME/I evaluation indicated that three of the six indicator's intercepts (Numeration, Geometry, Subtraction) were invariant in the joint group context. Establishing intercept invariance (albeit, partial invariance) indicates that a unit change in the latent construct is associated with comparable changes in the invariant indicator between groups across time. Further, intercept invariance establishes that longitudinal change can be attributed to true change in mathematics achievement and not changes related to measurement of the conceptual construct. Thus, the Numeration, Geometry, and Subtraction subtests showed comparable temporal change between groups.

With respect to the three non-invariant intercepts, findings related to the Addition subtest may be in part be due to benefits derived from participating in a reading intervention. That is, reading intervention students may have benefitted from the sound-symbol associations they were learning as evidenced by the increased addition intercepts (of about 0.5 an item) characteristic of this group's trajectory (see Figure 7). In particular, participation in a reading intervention may have improved students' retrieval of information (e.g., counting knowledge, computational strategies, long-term memory representations of basic arithmetic facts) from semantic memory, which enables the development of more complex mathematical skills (Geary, 1993; Geary & Burlingham-Dubree, 1989; Kaye, 1986).

For the Measurement subtest, intercept non-invariance suggested that the mathematics group showed an advantage over the reading group at the pre-intervention time point. Despite this early group difference, achievement on the Measurement subtest is similar at the mid- and post-intervention time points. This finding may indicate that different students differ on the average latent score at the pre-intervention time point and that a pre-existing group difference existed despite randomization to study conditions. Because children were grouped in part according to reading achievement scores and then randomly assigned to intervention conditions, a language retrieval difference (i.e., rapid autonomized naming) between groups may be responsible for non-invariance that characterized the Measurement subtest as well as that exhibited by the Addition and Time/Money subtests.

In summary, some evidence of indicator non-invariance was identified; however, evaluation of ME/I between groups across the three time points established a strong, partially invariant model of mathematics achievement. Thus, the indicators in the present study (Numeration, Geometry, Addition, Subtraction, Measurement, Time/Money), reliably and validly assessed mathematics achievement in to groups of children with mild disabilities. Further, the establishment of longitudinal ME/I is crucial to evaluating temporal change in a construct; without longitudinal ME/I, inferences concerning longitudinal growth cannot be unambiguously interpreted (Brown, 2006; Horn & McArdle, 1992). Establishing that mathematics achievement in the present study was characterized by comparable psychometric properties (equivalent form, factor loadings and intercepts) between groups across time, satisfies necessary conditions for evaluation and inferences concerning group differences and longitudinal change meaningful (Bryne et al., 1989; Muthen & Christoffersson, 1981).

4.2 Response to Mathematics Intervention

In the present study, it was projected that latent growth curves would be utilized to examine mathematics achievement growth. However, the data for the mathematics intervention group suggested that growth in mathematics achievement was curvilinear. It was therefore decided that latent growth models would not be estimated for a few reasons. To begin with, forcing the present curvilinear mathematics achievement data to fit a linear function would be model misspecification and evidenced by poor model fit test statistics. Further, the present data were limited to three time points. Without having data from four or more time points, the nature of change (or shape) that can be modeled is limited (Little, 2005). For instance, with data from four time points, change could be modeled as a quadratic function. Finally, although growth curves could have been investigated more generally, it was important to stay with the unbiased latent markers for mathematics achievement as this statistical methodology represents a strength in the present study compared to those that rely on traditional methods.

Although, latent growth models were not estimated, latent growth was examined through difference testing of mean achievement at mid- and post-intervention. As with modeling latent growth curves, difference testing of latent mean achievement is advantageous. This is because random error variance is separated out of the latent construct. In doing so, the biasing effects of random measurement errors can be accounted for (Medsker, Williams, & Holahan, 1994) and the distorting effects of measurement error on parameter estimates are mitigated (Chan, 1998; Vandenberg & Lance, 2000). Consequently, mean achievement represented in a latent variable is free of error.

Examination of differences in latent means at mid- and post-intervention indicated that the students in the mathematics group evidenced a small advantage over their peers who participated in a reading intervention. Although group differences at each of the time points were not statistically significant, these differences are in latent score units and therefore, can be interpreted as standard units relative to mathematics achievement in the reading intervention group at the pre-intervention time point. Thus, the 0.137 and 0.105 differences between the mathematics and reading group, at mid- and post-intervention can be interpreted as a small treatment effect. In addition to these group differences, the reported growth parameters indicated that, on average, students in the mathematics group evidenced more growth (as defined by total change in latent score units) from the pre- to post-intervention time points (between group difference of 0.146 over the school year in favor of the mathematics group).

In comparing the treatment effects in the present study with those in the discussed literature (Van Luit & Naglieri, 1999; Van Luit & Schopman, 2000), the effect size for the present sample’s total mathematics achievement raw mean (and standard deviation) at the pre- and post-intervention time points, was computed as in the cited studies (i.e., using the Cohen’s *d* formula reported in Table 7). Although, the effect sizes reported in the present study are weaker in magnitude than those in Van Luit and Naglieri

(1999) and Van Luit and Schopman (2000), there are several explanations for this difference. The present study utilized a norm-referenced measure (the KeyMath-R) developed to assess mathematics skills across children in kindergarten through ninth grade. Consequently, the 18 or 24 items that characterize each subtest may lack sensitivity to measure subtle intervention effects in elementary aged children with mild disabilities. In contrast, Van Luit and Naglieri (1999) utilized parallel versions of a researcher designed measure that consisted of 40 items each (20 multiplication and 20 division). Consequently, it is not surprising that the effect sizes reported in Van Luit and Naglieri (1999) were stronger than in the present study as researcher designed measures are often more sensitive to treatment effects than norm-referenced measures (Berkeley, Scruggs, & Mastropieri, 2010). In Van Luit and Schopman (2000) parallel versions of a norm-referenced measure (Utrecht Test for Number Sense; Van Luit, Van de Rijt, & Pennings, 1994) that included 40 items measuring counting skills and mathematics prerequisites was utilized. This measure of early numeracy consisted of eight parts: concepts of comparison, classification, correspondence, seriation, counting skills (using numerals, synchronized and shortened counting, and resultative counting), and general understanding of number. Thus, in both of Van Luit's studies, the criterion measure, whether researcher designed or norm-referenced mapped directly onto the skills targeted through the employed mathematics interventions. As a consequence, the outcome measures used by Van Luit and colleagues were sensitive to change in the target mathematics domain. It should be remembered, that the goal of the larger projects that made the present study possible, was to evaluate reading development. Given this focus, experimenter designed measures were not created to capture subtle changes in, for example, students' arithmetic development. Had experimenter designed or curriculum-based criterion measures been employed, the effect size estimates in the present study would likely be stronger than those presently reported.

Another possible explanation for the differences in effects sizes reported in the present study compared to the cited literature, is that there were differences between the present sample and that of Van Luit and colleagues. Specifically, the mean IQ of the children with MMR in the experimental group and the comparison group (70 and 71, respectively) of Van Luit and Naglieri (1999) was substantially higher than that of the children in the present study (63.03). This difference is likely due to the criteria for MMR school placement, which at the time of their study, included, “intellectual functioning below average (IQ range = 55-80)” (p. 99). In the Van Luit and Schopman’s (2000) study, mean IQ was provided in terms of the experimental conditions, which combined children with MMR and LD. For the experimental and comparison groups, mean IQ (and standard deviation) was 74.9 (13.10) and 79.1 (14.30). For the present study, the mean IQ (and standard deviation) for the mathematics and reading intervention groups was 74.80 (16.90) and 70.59 (16.07), respectively. Thus, as a group, the average IQ for the participants in the present study and those in Van Luit and Schopman (2000) appear to be similar; however, there appears to be more variability around IQ in the present study and consequently, more participants with lower IQ scores than in Van Luit and Schopman (2000).

In addition to the previous explanations, differences in the effect sizes between the present study and those of Van Luit and colleagues may be attributed to the interventions employed. To begin with, the comparison group in the present study was active. That is, the students in the comparison group participated in a rigorous reading intervention condition that was the focus of the larger projects as opposed to business as usual. Consequently, participants in the reading intervention condition may have benefitted from and perhaps, transferred reading gains to the mathematics context. Evidence from early research supports this view. In Gilmary (1967) elementary aged children with learning disabilities participated in a six-week summer school program. One group (M_{IQ} = 92.42) received instruction in reading and arithmetic, while the other (M_{IQ} = 98.5) received instruction in arithmetic only. The results indicated that the former group significantly outperformed the latter and that when IQ was accounted for the group difference was even more pronounced. Although there are apparent limitations (e.g., lack of ME/I evaluation and not including details related to the type of learning disability evidenced by each child) in Gilmary (1967), the results suggest that the addition of the reading component provided a boost to arithmetic instruction resulting in improved arithmetic achievement.

Besides the active comparison group, another notable difference between the mathematics intervention employed in the present study and those in the studies by Van Luit and colleagues, is that the mathematics intervention in the present study taught children multiple strategies to facilitate the development of skills related to several areas of mathematics (e.g., place value, number families, arithmetic facts, number relationships, measurement, regrouping in column addition and subtraction, problems solving). In contrast, interventions in Van Luit and colleagues focused on a particular skill set. For instance, the intervention in Van Luit and Naglieri (1999) concentrated on helping children use the results of simple multiplication and division problems in more complex related problems, while in Van Luit and Schopman (2000) the intervention focused on facilitating early numeracy skills and in particular, counting skill development. Concentrating on a particular skill area may be advantageous; however, it may also be advantageous to include multiple skills within a mathematics intervention, skills that are reviewed daily. Perhaps over time, the latter could support flexibility in students' mathematical thinking and strategy use. Taken together, the results concerning intervention effects in the present study are promising and suggest that students with MIDs and students with RDs benefitted from the intervention program.

4.3 Language Concerns in the Measurement of Mathematics Achievement

Measuring mathematics achievement in children that have a high incidence of speech and language disabilities, which are commonly associated with MIDs (Abbeduto, 2003) and RDs (Fletcher et al., 2007), is difficult. Although norm-referenced measures of mathematics achievement such as the KeyMath-R do not require reading on the students' part and the writing demands are minimal, such measures depend on expressive language skills. That is, in order to demonstrate competence on items within such measures, students must understand the examiner provided verbal prompt and then provide an expressive (oral) response for the solution. Thus, such measures rely heavily on oral responses from students in order to evaluate a student's competence on a particular task (Barker, 2010; Iacono & Cupples, 2004). All six subtests of the KeyMath-R used in the present study heavily rely on students' oral responses. Further, written computation is only permitted on the Addition and Subtraction subtests beginning with item seven of each subtest. Additionally, examination of the descriptive statistics displayed in Tables 1, 2, and 3, indicate that the median score on the Addition and Subtraction subtests did not exceed six. Thus, half of the students in each intervention condition did not have the opportunity to solve written problems on these two subtests because they reached the ceiling (i.e., three consecutive incorrect items). Despite dependence on relying on students' oral responses to evaluate mathematics achievement, the influence of this characteristic is minimal in the present study since mathematics achievement was measured reliably and validly between both interventions groups across three time points. The results of this study therefore suggest that the KeyMath-R subtests, as presently utilized, provide a reliable and valid measurement of mathematics achievement for children with MIDs and children with RDs. However, it may be best to use a global indicator rather than a marker of change for specific interventions due to reduced sensitivity.

4.4 Instructional Implications

This study demonstrated that the structure of mathematics achievement is equivalent in children with MIDs and children with RDs and that both groups of children show similar, positive responses to mathematics intervention. The present results and those of other randomized control studies (i.e., Van Luit and colleagues) indicate that students with MIDs and students with LDs similarly benefit from effective mathematics interventions. Thus, although IQ distinguishes between students from each special population, differential instruction should not be provided on basis of IQ alone. Instead, effective instruction is likely characterized by instructional groupings based on students' present levels of performance and intervention (or curriculum) that is designed that addresses student learning characteristics. For instance, ongoing assessment of students' mathematics proficiency may suggest that students who were previously grouped together (due to their previous levels of performance) are responding at different rates to intervention. Differing intervention response rates, however, may be attributed to a number of causes (e.g., absenteeism, attention, motivation, etc.) aside from IQ. Regardless, occasional regrouping of students (as in multitiered interventions), based on their individual rates of mathematic skills development, may be beneficial (Fletcher et al., 2007). When regrouping is not a feasible option, differentiated instruction could provide accommodations to lower performing students, while higher performing students complete more challenging (or enriching) work. In doing so, mathematics achievement of lower and higher skilled students can be fostered to the greatest extent possible.

As mentioned, designing instruction (or intervention) to address student learning characteristics may promote mathematics achievement. Design features of the intervention utilized in the present study, CMC and Base Ten Blocks, likely addresses important learner characteristics. For instance, providing students with frequent opportunities to respond (and *requiring* students to respond) during instruction likely increases student attention; while teaching to mastery through appropriate diagnosis and remediation of errors, improves mathematics proficiency (see Stein et al., 1997). Additionally, both the present intervention and those employed in Van Luit and Schopman (2000) and Miller and Mercer (1993) utilized concrete (i.e., objects) and semiconcrete (i.e., pictures) representations of number. Using concrete and semiconcrete representations of number may improve children's arithmetic calculation skills by supporting the development of procedural skills. Further, through repeated practice of connecting objects and pictures with abstract numbers, children can strengthen their memory of associations between arithmetic facts and their solution. One likely result of repeated practice with concrete and semiconcrete representations of number is transitioning children from the use of counting based arithmetic procedures to retrieval from memory (Van Luit & Schopman, 2000). Moreover, as children internalize arithmetic facts, it is probable that their conceptual understanding for number improves. Finally, given the high incidence of speech difficulties evidenced by students with MIDs (Abbeduto, 2003) and students with RDs (Fletcher et al., 2007), it could be beneficial to provide children from these special populations with a multi-component (or integrated) intervention that supports language development while facilitating mathematics achievement. Children with language delays or deficits, in particular, may benefit from such an approach.

4.5 Limitations and Future Directions

The design of this study answered questions about the structure of mathematics achievement, its measurement and response to mathematics intervention in children with MIDs and children with RDs. Some limitations, however, should be considered. To begin with, the present data were collected through two separate randomized control reading efficacy projects. As a result, the number of intervention (or contact) hours differed across the two projects. Specifically, second through fifth graders with MIDs participated in up to 120 hours of a given intervention, while second and third graders with RDs participated in up to 70 hours of a given intervention. Had children in both of the larger studies received the same number of intervention hours and the group with RDs included fourth and fifth graders, it is possible that the results of ME/I analyses would have been different. Of the three ME/I equality tests (i.e., form, loadings, and intercepts), non-invariance most likely would have manifested as intercept non-invariance between children with MIDs and children with RDs due to group mean differences on the six KeyMath-R subtests.

In addition to the difference in intervention hours between the two larger projects, the focus on reading development led to more children with MID (about two-thirds) being assigned to a reading intervention condition than the mathematics intervention condition. As a consequence, the reading intervention group is represented by a greater proportion of children with MID compared to RD, while the mathematics intervention group is closer to an equal representation of students from both special populations. Subsequently, future research should systematically evaluate the structure of mathematics achievement and its measurement longitudinally between additional special populations of children as in the present study. However, careful attention should be given to ensure that students from the respective populations represent the same grade level(s) in school, the number of intervention hours is more similar between populations, and intervention groups are more equally represented by the target special populations compared to the present study.

In regard to response to intervention, the present study was unable to accurately model change through latent growth models because growth in the mathematics intervention group was curvilinear and because mathematics achievement data were not available from a fourth time point. Subsequently, the data were not forced to fit a linear growth model. Longitudinal work is expensive and time intensive; however, future research could benefit from collecting mathematics achievement data across four time points. In doing so, latent growth could be estimated using linear and curvilinear model specifications and the tenability of each model can be evaluated.

Finally, the two larger projects that made the present study possible, focused on reading development. As such, instructional grouping procedures were based in part on measures of students' reading achievement. In some instances, this may have resulted in intervention groups that were heterogeneous with respect to students' mathematics skills. Consequently, it may have been difficult to maximize learning for *all* students within an overly heterogeneous group and may have mitigated learning in a small number of cases. It would be beneficial for future studies to group students with respect to their arithmetic skill development as mastery of basic arithmetic skills is crucial to the development of more complex mathematical skills (Geary, 1993; Geary & Burlingham-Dubree, 1989; Kaye, 1986).

4.6 Conclusion

In conclusion, the findings from the present study indicate that the form of the latent mathematics achievement factor was unidimensional and the pattern of indicator-factor loadings between groups across time are equivalent. Because of the longitudinal nature of this study, it can be concluded that the nature (or structure) of mathematics achievement in children with MIDs and children with RDs is fundamentally the same and temporally stable. The results therefore support the assumption that students with MIDs and students with RDs move through similar, if not the same, steps as typically developing children in acquiring mathematics proficiency. In addition, despite a few sources of non-invariance, the present results indicate that the measurement of mathematics achievement was equivalent between intervention groups and consequently, populations across time. Because equivalent form and measurement between groups across time was established, prerequisites to evaluating group differences and change, students' response to a mathematics intervention was evaluated. It was demonstrated that students randomly assigned to a mathematics intervention condition evidenced a small advantage over students randomly assigned to a reading intervention condition with respect to latent mathematics achievement at the mid- and post-intervention time points. Evidence also was provided that indicated that students in the former group displayed more growth than that shown by the latter group. Thus, the findings suggest that students with MID and students with RD benefited from the same mathematics intervention; however, the treatment effects were small.

REFERENCES

Abbeduto, L. (2003). *International review of research in mental retardation: Language and communication*. New York: Academic Press.

Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *The Journal of Human Resources, 30*(3), 409-438.

American Association on Intellectual and Developmental Disabilities. (2010). Definition of intellectual disability. Retrieved from http://www.aaidd.org/content_100.cfm

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Barker, R. M. (2010). Modeling phonological processing for children with mild intellectual disabilities: The relationship between underlying phonological abilities and associated language variables. *Psychology Dissertations.* Paper 77.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42,* 825-829. doi:10.1016/j.paid.2006.09.024

Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995-2006: A meta-analysis. *Remedial And Special Education, 31*(6), 423-436. doi:10.1177/0741932509355988

Boomsma, A., & Hoogland, J. J. (2001). *The robustness of LISREL modeling revisited.* Amsterdam: Sociometric Research Foundation.

Brankaer, C., Ghesquière, P., & De Smedt, B. (2011). Numerical magnitude processing in children with mild intellectual disabilities. *Research in Developmental Disabilities, 32*, 2853-2859. doi:10.1016/j.ridd.2011.05.020

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Cameto, R. (2005). Employment of youth with disabilities after high school. In M. Wagner, L. Newman, R. Cameto, N. Garza, & P. Levine, (Eds.), *After high school: A first look at the postschool experiences of youths with disabilities. A report from the National Longitudinal Transition Study-2 (NLTS-2).* Menlo Park, CA: SRI International. Retrieved April 15, 2013, from http://www.nlts2.org/reports/2005_04/nlts2_report_2005_04_ch5.pdf

Carnine, D. (1991). Curricular interventions for teaching higher order thinking to all students: Introduction to the special series. *Journal of Learning Disabilities, 24*(5), 261-269. doi:10.1177/002221949102400502

Carnine, D. (1997). Instructional design in mathematics for students with learning disabilities. *Journal of Learning Disabilities, 30*, 130–141. doi:10.1177/002221949703000201

Cawley, J. F., Fitzmaurice-Hayes, A., & Shaw, R. (1988). *Mathematics for the mildly handicapped—A guide to curriculum and instruction.* In S. Miller, & C. D. Mercer (1993). Using data to learn concrete-semiconcrete-abstract instruction for students with math disabilities. *Learning Disabilities Research and Practice, 8*(2) (p. 89).

Cawley, J. F., & Miller, J. H. (1989). Cross-sectional comparisons of the mathematical performance of children with learning disabilities: Are we on the right track toward comprehensive programming? *Journal of Learning Disabilities, 22*, 250-259.

Cawley, J. F., Miller, J. H., & School, B. A. (1987). A brief inquiry of arithmetic word-problem-solving among learning disabled secondary students. *Learning Disabilities Focus, 2*(2), 87-93.

Cawley, J. F., Parmar, R., Foley, T. E., Salmon, S., & Roy, S. (2001). Arithmetic performance of students: Implications for standards and programming. *Exceptional Children*, *67*(3), 311-328.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structure analysis (LMACS) and multiple indicator latent growth modeling. *Organizational Research Methods 1*(4), 421-483.

Connolly, A. (1988). *Key Math-Revised: A Diagnostic Inventory of Essential Mathematics*. Circle Pines, MN: American Guidance Service.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*(2), 81-100. doi:10.1037/a0015914

Engelmann, S., & Becker, W. (1995). *Connecting math concepts: Level K.* Eugene, OR: Engelmann/Becker Corporation.

Engelmann, S., & Carnine, D. (1976). *SRA distar arithmetic II: Teacher's guide.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., & Carnine, D. (1992-2003). *SRA connecting math concepts: Level A.* Columbus, OH: SRA/McGraw-Hill.

Ferrari, M. (2009). Borderline intellectual functioning and the intellectual disability construct. *Intellectual and Developmental Disabilities*, *47*(5), 386-389. doi:10.1352/1934-9556-47.5.386

Field, A. P. (2012). Exploring data: The beast of bias. Retrieved from www.statisticshell.com/docs/exploringdata.pdf

Fletcher, J. M., Lyon, R. G., Fuchs, L. S., & Barnes, M. (2007). *Learning disabilities: From identification to intervention.* New York, NY: The Guilford Press.

Foster, M. E., Sevcik, R. A., Romski, M., & Morris, R. D. (2014). The effects of phonological awareness and naming speed on mathematics skills in children with mild intellectual disability. *Developmental Neurorehabilitation*. Advance online publication. doi:10.3109/17518423.2013.843603

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37*, 293-306.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345-362. doi:10.1037/0033-2909.114.2.345

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23-27. doi:10.1177/0963721412469398

Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology, 47*,175-192. doi:10.1016/0022-0965(89)90028-3

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, *77*(3), 236-263. doi:10.1006/jecp.2000.2561

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

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

Gilmary, S. (1967). Transfer effects of reading remediation to arithmetic computation when intelligence is controlled and all other school factors eliminated. *The Arithmetic Teacher, 14*, 17-20.

Glass, R. M., Christiansen, J., & Christiansen, J. L. (1982). Teaching exceptional students in the classroom. In K. M. McCoy, *Teaching special learners: In the general education classroom* (2nd ed., p.44).

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *The Journal of Applied Behavioral Science, 12*(2), 133-*157.*

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

Heikura, U., Linna, S. L., Olsén, P., Hartikainen, A. L., Taanila, A., & Järvelin, M.R. (2005). Etiological survey on intellectual disability in the Northern Finland Birth Cohort 1986. *American Journal on Mental Retardation, 110*(3), 171-180.

Hoard, M. K., Geary, D. C., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Performance of low- and average-IQ children. *Mathematical Cognition, 5*, 65-91.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-143.

Iacono, T. A., & Cupples, L. (2004). Assessment of phonemic awareness and word reading skills of people with complex communication needs. *Journal of Speech, Language, and Hearing Research, 47*, 437-449.

Individuals with Disabilities Education Act of 2004. 20 U.S.C.§§ Section 1401 (30). Retrieved from http://idea.ed.gov/explore/view/p/%2Croot%2Cregs%2C300%2CA%2C300%252E8%2C

Joensen, J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *The Journal of Human Resources, 44*(1), 171-198.

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33*(6), 567-578. doi:10.1177/002221940003300605

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850.

Jordon, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94*(3), 586-597. doi:10.1037/0022-0663.94.3.586

Kaufman, A. S., & Kaufman, A. L. (1990). *Kaufman-Brief Intelligence Test*. Circle Pines, MN: American

Guidance Service.

Kaye, D. B. (1986). The development of mathematical cognition. *Cognitive Development, 1*(2), 157-170. doi:10.1016/S0885-2014(86)80017-X

Kelly, B., Gersten, R., & Carnine, D. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students and high school students with learning disabilities. *Journal of Learning Disabilities, 23*(1), 23-29. doi:10.1177/002221949002300108

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford.

Kroesbergen, E. H., & Van Luit, J. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial And Special Education, 24*(2), 97-114. doi:10.1177/07419325030240020501

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *The Journal of Child Psychology and Psychiatry, 51*(3), 287-294. doi:10.1111/j.1469-7610.2009.02164.x

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek, & T. D. Little (2006). *Handbook of personality development*. Mahwah, N.J.: Lawrence Erlbaum.

Mattingly, J. C., & Bott, D. A. (1990). Teaching multiplication facts to students with learning problems. *Exceptional Children, 56*(5), 438-449.

McArdle, J. J. (1996). Current directions in structural factor analysis. *Current Directions in Psychological Science, 5*(1), 11-18.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum. In T. A. Brown (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

McKenzie, M. A., Marchand-Martella, N. E., Moore, M. E., & Martella, R. C. (2005). Teaching basic math skills to preschoolers using Connecting Math Concepts Level K. *Journal of Direct Instruction, 4*(1), 85-94.

McLean, P., Laycock, M., & Smart, M. (1990). *Building understanding with base ten blocks.* Hayward, CA: Activity Resources Company, Incorporated.

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, *20*(2), 439-464.

Miller, S. P., & Mercer, C. D. (1993). Using data to learn concrete-semiconcrete-abstract instruction for students with math disabilities. *Learning Disabilities Research and Practice, 8*(2), 89-96.

Miller, S. P., & Mercer, C. D. (1997). Educational aspects of mathematics disabilities. *Journal of Learning Disabilities, 30*(1), 47-56. doi:10.1177/002221949703000104

Miller, S. P., Mercer, C. D., & Dillon, A. S. (1992). CSA: Acquiring and retaining math skills. *Intervention In School and Clinic, 28*(2), 105-110. doi: 10.1177/105345129202800206

Miller, C. J., Sanchez, J., & Hynd, G. W. (2003). Neurological correlates of reading disabilities. In H. Swanson, K.R. Harris, S. Graham (Eds.), *Handbook of learning disabilities* (pp. 242-255). New York: Guilford Press.

Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares and J. J. McArdle (2013). *Contemporary psychometrics* (pp. 153-171). New York: Psychology Press.

Morris, R. D. (2007). Treatment of developmental reading disabilities. Georgia State University: National Institutes of Health.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, *46*(4), 407-419.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards (Mathematics).* Washington D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Joint Committee on Learning Disabilities. (1991). Learning disabilities: Issues on definition. *American Speech, and Hearing Association, 33*, (Suppl. 5), 18–20.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Retrieved from the U.S. Department of Education Web site: http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J. & Svinicki, J. (1984). *Battelle Developmental Inventory*. Itasca, IL: Riverside.

Obi, O., Van Naarden Braun, K., Baio, J., Drews-Botsch, C., Devine, O., & Yeargin-Allsopp, M. (2011). Effect of Incorporating Adaptive Functioning Scores on the Prevalence of Intellectual Disability. *American Association on Intellectual and Developmental Disabilities, 116*(5), 360-370. doi:10.1352/1944-7558-116.5.360

Parmar, R. S., Cawley, J. F., & Miller, H. (1994). Differences in mathematics performance between students with learning disabilities and students with mild retardation. *Exceptional Children, 60*, 549-564.

Pierangelo, R., & Giuliani, G. A. (2007). *EDM: The educator's diagnostic manual of disabilities and disorders* (pp. 1-52). San Francisco, CA: Jossey-Bass.

Podell, D. M., Tournaki-Rein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted instruction among students with mild mental handicaps. *Education and Training in Mental Retardation, 27*(3), 200-206.

Polloway, E. A., Lubin, J., Smith, J. D., & Patton, J. R. (2010). Mild intellectual disabilities: Legacies and trends in concepts and educational practices. *Education and Training in Autism and Developmental Disabilities, 45*(1), 54-68.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., & Trautmann, U. (n.d). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics Part A, 140A*(19), 2063-2074. doi:10.1002/ajmg.a

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

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics*, *86*(2), 497-513.

Sadler, P. M., & Tai, R. H. (2007). The two high school pillars supporting college science. *Science, 317*, 457-458.

Schalock, R. (2011). The evolving understanding of the construct of intellectual disability. *Journal of Intellectual and Developmental Disability*, *36*(4), 227-237. doi:10.3109/13668250.2011.624087

Sevcik, R. A. (2005). Evaluating the effectiveness of reading interventions for students with mild mental retardation. Georgia State University: U.S. Department of Education.

Simos, P. G., Breier, J. I., Fletcher, J. M., Bergman, E. E., & Papanicolaou, A. C. (2000). Cerebral mechanisms involved in word reading in dyslexic children: A magnetic source imaging approach *Cerebral Cortex*, *10*(8), 809-816. doi:10.1093/cercor/10.8.809

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173-180.

Stein, M., Silbert, J., & Carnine, D. (1997). *Designing effective mathematics instruction: A direct instruction approach.* Upper Saddle River, NJ: Prentice Hall. In A. M. Przychodzin, N. E. Marchand-Martella, R. C. Martella, & D. Azim (2004). Direct instruction mathematics programs: An overview and research summary. *Journal of Direct Instruction, 4*(1), 61-64.

The President's Commission on Excellence in Special Education. (2002, July). *A new era: Revitalizing special education for children and their families.* Washington, DC: U.S. Department of Education.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(4), 4-70. doi: 10.1177/109442810031002

Van Luit, J. H. (1987). Teaching impulsive children with arithmetic deficits in special education: A self-instructional training program. *European Journal of Special Needs Education, 2*, 237–246.

Van Luit, J. H. (1994). The effectiveness of structural and realistic arithmetic curricula in children with special needs. *European Journal of Special Needs Education*, *9*, 16–26.

Van Luit, J. H., & Naglieri, J. A. (1999). Effectiveness of the MASTER program for teaching special children multiplication and division. *Journal of Learning Disabilities*, *32*(2), 98-107. doi:10.1177/002221949903200201

Van Luit, J. H., & Schopman, E. M. (2000). Improving early numeracy of young children with special educational needs. *Remedial and Special Education*, *21*(1), 27-40. doi:10.1177/074193250002100105

Van Luit, J. H., Van de Rijt, B. M., & Pennings, A. H. (1994). *Utrechtse Getalbegrip Toets* [*Utrecht Test of Number Sense*]. Doetinchem, The Netherlands: Graviant.

Warner, M., Alley, G., Schumaker, J., Deshler, D., & Clark, F. (1980). *An epidemiological study of learning disabled adolescents in secondary school: Achievement and Ability, socioeconomic status, and school experiences.* Research Report 13. Institute for Research in Learning Disabilities: University of Kansas.

Whitley, B. E. (2002). *Principles of research in behavioral science* (2nd ed.). New York: McGraw-Hill.

Wilson, C. L., & Sindelar, P. T. (1991). Direct instruction in math word problems: Students with learning disabilities. *Exceptional Children*, *57*(6), 512-519.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests-Revised.* Circle Pines, MN: American Guidance Service.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structural analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165-200.

APPENDICES

Appendix 1

Descriptive Statistics by Disability Group: Pre-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
MID only									
NUM	265	4.99	2.69	4.0	0	13	3	1.01 (0.15)	0.98 (0.30)
GEO	265	3.69	3.09	4.0	0	14	51	0.60 (0.15)	-0.30 (0.30)
ADD	265	3.15	2.76	2.0	0	14	34	1.25 (0.15)	1.30 (0.30)
SUB	265	1.22	1.57	1.0	0	10	115	1.74 (0.15)	3.96 (0.30)
MST	265	2.63	2.29	2.0	0	13	39	1.54 (0.15)	2.86 (0.30)
TIMO	265	2.20	2.25	2.0	0	12	60	1.62 (0.15)	3.17 (0.30)
RD only									
NUM	132	8.34	3.00	8.0	3	19	2	0.74 (0.21)	0.74 (0.42)
GEO	131	8.34	3.29	8.0	0	15	1	-0.23 (0.21)	-0.50 (0.42)
ADD	132	6.13	2.36	6.0	1	13	2	0.45 (0.21)	0.35 (0.42)
SUB	132	3.52	1.95	3.0	0	9	1	0.53 (0.21)	-0.37 (0.42)
MST	131	7.37	2.82	8.0	1	12	3	-0.29 (0.21)	-0.75 (0.42)
TIMO	131	3.96	2.56	4.0	0	12	12	0.65 (0.21)	0.36 (0.42)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 1

Descriptive Statistics by Disability Group: Mid-point

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
MID only									
NUM	253	5.84	2.94	5.0	0	15	1	0.97 (0.15)	0.65 (0.31)
GEO	253	4.70	3.37	5.0	0	13	31	0.33 (0.15)	-0.75 (0.31)
ADD	253	4.37	3.22	3.0	0	14	19	0.77 (0.15)	-0.07 (0.31)
SUB	253	1.56	1.85	1.0	0	10	90	1.62 (0.15)	2.96 (0.31)
MST	253	3.21	2.69	3.0	0	12	25	1.15 (0.15)	0.59 (0.31)
TIMO	253	2.81	2.52	2.0	0	13	36	1.26 (0.15)	1.35 (0.31)
RD only									
NUM	132	10.16	3.11	10	4	19	0	0.21 (0.21)	-0.47 (0.42)
GEO	132	10.16	3.30	10.5	2	16	0	-0.36 (0.21)	-0.59 (0.42)
ADD	133	7.31	2.47	7.0	1	13	0	-0.15 (0.21)	-0.52 (0.42)
SUB	132	4.48	2.23	4.0	1	11	0	0.46 (0.21)	-0.35 (0.42)
MST	132	8.24	2.83	9.0	1	16	0	-0.41 (0.21)	0.28 (0.42)
TIMO	132	5.52	2.86	5.0	1	16	0	1.13 (0.21)	2.36 (0.42)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 1

Descriptive Statistics by Disability Group: Post-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
MID only									
NUM	245	6.38	3.04	6.0	0	15	1	0.72 (0.16)	0.07 (0.31)
GEO	245	5.42	3.33	5.0	0	15	18	0.27 (0.16)	-0.32 (0.31)
ADD	245	5.03	3.51	5.0	0	14	15	0.57 (0.16)	-0.47 (0.31)
SUB	245	2.05	2.06	1.0	0	10	61	1.09 (0.16)	0.64 (0.31)
MST	245	3.66	2.83	3.0	0	14	23	0.96 (0.16)	0.37 (0.31)
TIMO	245	3.44	2.81	3.0	0	14	26	1.10 (0.16)	1.01 (0.31)
RD only									
NUM	136	10.96	3.24	11.0	5	20	0	0.24 (0.21)	-0.42 (0.41)
GEO	136	10.96	3.02	11.0	4	17	0	-0.17 (0.21)	-0.78 (0.41)
ADD	136	8.07	2.52	8.0	2	14	0	0.05 (0.21)	-0.58 (0.41)
SUB	136	4.90	2.47	4.5	0	11	3	0.41 (0.21)	-0.33 (0.41)
MST	136	8.88	2.99	9.0	1	16	0	0.42 (0.21)	0.48 (0.41)
TIMO	136	6.55	3.56	6.0	0	21	1	1.40 (0.21)	2.72 (0.41)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. MID = Mild intellectual disability, RD = Reading Disability. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 2

Descriptive Statistics for Combined Disability Groups: Pre-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
NUM	397	6.10	3.21	5.0	0	19	3	0.80 (0.12)	0.51 (0.24)
GEO	396	5.23	3.84	5.0	0	15	52	0.36 (0.12)	-0.76 (0.25)
ADD	397	4.14	2.98	4.0	0	14	34	0.65 (0.12)	-0.16 (0.24)
SUB	397	1.99	2.01	1.0	0	10	116	1.06 (0.12)	0.63 (0.24)
MST	396	4.20	3.34	3.0	0	13	39	0.72 (0.12)	-0.59 (0.25)
TIMO	396	2.78	2.51	2.0	0	12	72	1.16 (0.12)	1.28 (0.25)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 2

Descriptive Statistics for Combined Disability Groups: Mid-point

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
NUM	385	7.32	3.63	7.0	0	19	1	0.56 (0.12)	-0.43 (0.25)
GEO	385	6.57	4.23	6.0	0	16	31	0.18 (0.12)	-0.90 (0.25)
ADD	386	5.38	3.29	5.0	0	14	19	0.26 (0.12)	-0.82 (0.25)
SUB	385	2.56	2.43	2.0	0	11	90	0.95 (0.12)	0.25 (0.25)
MST	385	4.93	3.64	3.0	0	16	25	0.48 (0.12)	-0.92 (0.25)
TIMO	385	3.74	2.93	3.0	0	16	36	1.06 (0.12)	1.43 (0.25)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 2

Descriptive Statistics for Combined Groups: Post-intervention

Variable	*N*	*M*	*SD*	*Median*	*Min*	*Max*	#0's	*Skew (SE)*	*Kurtosis (SE)*
NUM	381	8.01	3.81	7.0	0	20	1	0.47 (0.13)	-0.43 (0.25)
GEO	381	7.39	4.17	7.0	0	17	18	0.12 (0.13)	-0.74 (0.25)
ADD	381	6.12	3.51	6.0	0	14	15	0.13 (0.13)	-0.83 (0.25)
SUB	381	3.07	2.60	3.0	0	11	64	0.77 (0.13)	-0.07 (0.25)
MST	381	5.52	3.82	5.0	0	16	23	0.41 (0.13)	-0.86 (0.25)
TIMO	381	4.55	3.44	4.0	0	21	27	1.24 (0.13)	2.34 (0.25)

Note. *N* = Number of participants. *M* = Mean, *SD* = Standard deviation, #0's = Number of scores of 0, *SE* = Standard error. NUM = Numeration, GEO = Geometry, ADD = Addition, SUB = Subtraction, MST = Measurement, TIMO = Time/Money.

Appendix 3

Within Group Longitudinal Measurement Invariance Evaluation: Reading Group

	χ^2_{SB} (*df*)	χ^2_{SBdiff}	Δdf	CFI	RMSEA (90% CI)	$P_{RMSEA} \leq .05$	SRMR
Step 1. Equal Form	175.05 (114)***			.984	.046 (.032, .059)	.67	.028
Step 2. Equal Factor Loadings	183.85 (124)***	8.29	10	.984	.044 (.030, .057)	.77	.031
Step 3. Equal Intercepts	**210.69 (134)*****	**26.31****	**10**	**.980**	**.048 (.035, .063)**	**.61**	**.039**
Step 4. Equal Intercept: Numeration	183.85 (124)***			.984	.044 (.030, .057)	.78	.031
Step5. Equal Intercept: Geometry	**190.07 (126)*****	**6.37***	**2**	**.983**	**.045 (.031, .058)**	**.73**	**.033**
Step 6. Equal Intercept: Addition	**193.07 (126)*****	**8.96***	**2**	**.982**	**.046 (.032, .058)**	**.69**	**.035**
Step 7. Equal Intercept: Subtraction	186.05 (126)***	1.90	2	.984	.043 (.029, .056)	.79	.031
Step 8. Equal Intercept: Measurement	187.87 (128)***	1.77	2	.984	.043 (.029, .056)	.80	.032
Step 9. Equal Intercept: Time/Money	192.61 (130)***	4.64	2	.984	.044 (.030, .056)	.78	.034

Note. χ^2_{SB} = Satorra-Bentler scaled χ^2; χ^2_{SBdiff} = difference test; Δdf = change in degrees of freedom; CFI = Comparative fit index, RMSEA = Root mean square error of approximation, $P_{RMSEA} \leq .05$ = Test of close fit, SRMR = Standardized root mean square residual. Bold identifies non-invariance via difference testing.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Appendix 3

Within Group Longitudinal Measurement Invariance Evaluation: Mathematics Group

	χ^2_{SB} (*df*)	χ^2_{SBdiff}	Δdf	CFI	RMSEA (90% CI)	$P_{RMSEA} \leq .05$	SRMR
Step 1. Equal Form	183.46 (114)***			.976	.064 (.046, .081)	.09	.026
Step 2. Equal Factor Loadings	**220.31 (124)*****	**39.44*****	**10**	**.967**	**.072 (.057, .088)**	**.01**	**.043**
Step 3. Equal Factor Loadings: Numeration	183.46 (114)***	0	0	.976	.064 (.046, .081)	.09	.026
Step 4. Equal Factor Loadings: Geometry	184.89 (116)***	1.57	2	.976	.063 (.046, .080)	.10	.026
Step5. Equal Factor Loadings: Addition	185.12 (118)***	.19	2	.977	.062 (.044, .079)	.13	.027
Step 6. Equal Factor Loadings: Subtraction	189.44 (120)***	5.53	2	.976	.063 (.045, .079)	.11	.030
Step 7. Equal Factor Loadings: Measurement	192.34 (122)***	2.55	2	.976	.062 (.045, .079)	.11	.031
Step 8. Equal Factor Loadings: Time/Money	**220.31 (124)*****	**39.62*****	**2**	**.967**	**.072 (.057, .088)**	**.01**	**.043**
Step 9. Equal Intercepts	**220.76 (130)*****	**204.79*****	**2**	**.969**	**.069 (.053, .084)**	**.03**	**.043**
Step 10. Equal Intercept: Numeration	192.34 (122)***	0	0	.976	.062 (.045, .079)	.114	.031
Step 11. Equal Intercept: Geometry	195.88 (124)***	3.49	2	.975	.063 (.045, .079)	.11	.031
Step 12. Equal Intercept: Addition	196.94 (126)***	.99	2	.976	.062 (.044, .078)	.13	.031
Step 13. Equal Intercept: Subtraction	198.79 (128)***	1.94	2	.976	.061 (.044, .077)	.14	.031
Step 14. Equal Intercept: Measurement	**220.76 (130)*****	**17.31*****	**2**	**.969**	**.069 (.053, .084)**	**.03**	**.041**

Note. χ^2_{SB} = Satorra-Bentler scaled χ^2; χ^2_{SBdiff} = difference test; Δdf = change in degrees of freedom; CFI = Comparative fit index, RMSEA = Root mean square error of approximation, $P_{RMSEA} \leq .05$ = Test of close fit, SRMR = Standardized root mean square residual. Bold identifies non-invariance via difference testing.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Appendix 4

Unstandardized Latent and Residual Correlations (with Standard Deviations): Reading Group

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. MATH1	--											
2. MATH2	1.09 (0.03)***	--										
3. MATH3	1.16 (0.03)***	1.31 (0.07)***	--									
4. NUM1				--								
5. NUM2				0.58 (0.19)**	--							
6. NUM3				0.19 (0.26)	0.19 (0.23)**	--						
7. GEO1							--					
8. GEO2							1.61 (0.56)**	--				
9. GEO3							1.57 (0.48)**	1.60 (0.48)**	--			
10. ADD1										--		
11. ADD2										1.29 (0.33)***	--	
12. ADD3										1.56 (0.38)***	2.21 (0.34)***	--

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; MATH = Mathematics achievement latent factor; NUM = Numeration, GEO = Geometry, ADD = Addition. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point.

Appendix 4

Unstandardized Residual Correlations (with Standard Deviations): Reading Group

	13.	14.	15.	16.	17.	18.	19.	20.	21.
13. SUB1	--								
14. SUB2	0.38 (0.17)*	--							
15. SUB3	0.28 (0.17)	0.37 (0.20)	--						
16. MST1				--					
17. MST2				1.20 (0.33)***	--				
18. MST3				1.22 (0.31)***	1.70 (0.30)***	--			
19. TIMO1							--		
20. TIMO2							1.02 (0.24)***	--	
21. TIMO3							1.22 (0.24)***	1.82 (0.31)***	--

Note. *p < .05, **p < .01, ***p < .001; SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point.

Appendix 4

Unstandardized Latent and Residual Correlations (with Standard Deviations): Mathematics Group

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. MATH1	--											
2. MATH2	1.25 (0.17)***	--										
3. MATH3	1.28 (0.18)***	1.50 (0.20)***	--									
4. NUM1				--								
5. NUM2				0.38 (0.17)	--							
6. NUM3				1.00 (0.28)***	0.69 (0.22)**	--						
7. GEO1							--					
8. GEO2							1.72 (0.65)**	--				
9. GEO3							1.03 (0.54)	2.01 (0.54)***	--			
10. ADD1										--		
11. ADD2										0.22 (0.20)	--	
12. ADD3										0.38 (0.26)	1.31 (0.28)	--

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; MATH = Mathematics achievement latent factor; NUM = Numeration, GEO = Geometry, ADD = Addition. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point.

Appendix 4

Unstandardized Residual Correlations (with Standard Deviations): Mathematics Group

	13.	14.	15.	16.	17.	18.	19.	20.	21.
13. SUB1	--								
14. SUB2	0.23 (0.13)	--							
15. SUB3	0.23 (0.17)	0.35 (0.20)	--						
16. MST1				--					
17. MST2				1.85 (0.35)**	--				
18. MST3				1.78 (0.39)***	2.51 (0.48)***	--			
19. TIMO1							--		
20. TIMO2							0.40 (0.19)*	--	
21. TIMO3							0.64 (0.26)*	0.98 (0.44)*	

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; SUB = Subtraction, MST = Measurement, TIMO = Time/Money. 1 = Pre-intervention time point, 2 = Mid-intervention time point, 3 = Post-intervention time point.