2019

The Effect of STEM and non-STEM Education on Student Mathematics Ability in Third Grade

Elke Hyacinth
Walden University

<u>Walden University</u>

College of Education

This is to certify that the doctoral dissertation by

Elke Jessonya Hyacinth

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Patricia Anderson, Committee Chairperson, Education Faculty
Dr. Amy White, Committee Member, Education Faculty
Dr. Jennifer Lapin, University Reviewer, Education Faculty

The Office of the Provost

Walden University
2019

Abstract

The Effect of STEM and non-STEM Education on Student Mathematics Ability in Third Grade

by

Elke Jessonya Hyacinth

MA, Walden University, 2006

BS, Prairie View A&M University, 1986

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Curriculum, Instruction, and Assessment

Walden University

November, 2019

Abstract

Although early mathematics instruction is predictive of future mathematics achievement, the effects of STEM-based mathematics instruction on mathematics gains in elementary school have been largely unexplored. The purpose of this quantitative study was to determine whether mathematics scores from third grade student state-mandated standardized mathematics test differ between students who were enrolled in STEM schools and students who were enrolled in non-STEM schools in the largest school district located in a Southwestern state in the United States. Polya's problem-solving heuristics formed the theoretical framework because of their relevance to concepts on the third grade mathematics test. Two research questions focused on intraindividual changes and interindividual changes over time in standardized mathematics test scores of third grade students who were enrolled in 18 STEM and 18 non-STEM schools. Analyses included growth curve modeling and a one-way random effect ANOVA to determine individual growth trajectories of mathematics test scores from individual schools over time from 2012 through 2017. The results indicated that there were no intraindividual differences in growth over time within schools, and there were interindividual changes in growth over time between schools, but the changes could not be explained by the independent variables, STEM and non-STEM schools. Findings were not consistent with the literature, which indicated early STEM-based mathematics instruction is more beneficial than traditional instruction. This study offers implications for positive social change by demonstrating equivalent results of STEM to non-STEM instruction, which may encourage more hands-on, inquiry-based learning for all children.

The Effect of STEM and non-STEM Education on Student Mathematics Ability in Third Grade

by

Elke Jessonya Hyacinth

MA, Walden University, 2006

BS, Prairie View A&M University, 1986

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Curriculum, Instruction, and Assessment

Walden University

November, 2019

Dedication

I dedicate this paper to my wonderful parents, Napoleon Theophilus and Bobbie Jean Milton. I am forever grateful to you for teaching me the value of lifelong learning. To my beautiful daughter, Mayowa Cherokee Milton, thank you for your unconditional love for me through this process. As a single parent, countless mornings, afternoons, and evenings of toiling through theory and writing meant that I was unavailable for several outings that were important to you. You know that I love you madly, I truly appreciate your sacrifice and unselfishness, and I hope you can find it in your heart to forgive me. To my siblings who encouraged me at every turn; you are the greatest family ever. John, my husband, thank you for always listening and knowing exactly what to say to help me process, stay calm, and pushing me towards the end. I also want to thank my step-daughter, Ladedra. Talking my ears off even as I struggled to write this paper and have to listen simultaneously was a challenge, but your persistence reminded about the importance of family, so thank you for keeping me grounded.

Acknowledgments

Soon after I began this journey, one obstacle after another presented itself, and if not for the favor and grace and mercies of the Lord, and the prayers of many, I would not have survived. I know I would not have arrived here without the guidance and support of my wonderful committee. First, thank you to my Chair, Dr. Patricia Nan Anderson. Your mentorship, your enduring support, our countless conference calls, your knowledge about APA style, STEM education, early childhood education, and a long list of related topics guided me through this process. I will be forever grateful to you for your tireless work with me through this colossal endeavor, which has taught me to be a scholar. Your passion for your work has inspired me and I am looking forward to making positive social change in my own line of work in STEM education. The balance of your patience and ability to push and be firm with me has prepared me to take on the world! Dr. White, thank you for your expertise as my methodologist. Your knowledge about statistics, your understanding about this scary process, and your compassion when I was in tears comforted me during the times I really needed it, and encouraged me to keep going. Thank you to all of my Walden professors who shared information and provided opportunities with me along the way. I want to send a special thank you to Megan Herrley, my hands-on advisor, for providing assistance along the way. Your check-ins and check-ups meant a great deal to me. Finally, I must thank my circle of friends and colleagues for cheering me on, especially during the difficult times.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

In this study, I examined the difference on state-mandated standardized mathematics test scores between third grade students who were enrolled in science, technology, engineering, and mathematics (STEM) schools and third grade students who were not enrolled in STEM schools. STEM support is more beneficial when introduced during early childhood education, and the level of mathematical skill gained in preschool is predictive of mathematics achievement throughout high school (Clements & Sarama; 2016; McClure et al., 2017; Oberle, Schonert-Reichl, Hertzman, & Zumbo, 2014; Rosicka, 2016). Despite growing evidence about the advantages of early mathematics instruction, virtually no research exists that compares the effect of STEM and non-STEM, or traditional education, on third grade student performance on high- or low-stakes standardized mathematics tests (Chiu, Price, & Ovrahim, 2015; Clements & Sarama, 2016; Ejiwale, 2013; McClure et al., 2017).

The results of this research provided information to advance positive social change by clarifying the importance of early-grade STEM pedagogy in supporting mathematics achievement, especially because third grade students are now subject to standardized testing. The results of this study can help educational stakeholders in educational planning for early-grade instruction. In addition, this study's results might catalyze changes in local, state, or national educational decisions, including policies and funding that influence endorsement of high quality STEM programs and curricula in elementary education, which is partly provided by an arm of the federal government, the

Committee on Science, Technology, Engineering and Math Education (Lamberg & Trzynadlowski, 2015).

This chapter includes a brief review of the background and outlook on STEM education in the primary grades and early mathematics instruction and provides a description of the problem and purpose of this study, its theoretical foundation, and the research questions (RQs) that guided data collection. I also address the limitations, assumptions, scope and delimitations of this study and identify the steps taken to ensure ethical treatment of data and stakeholders.

Background

A survey of educational stakeholders on trends regarding STEM education across all grade levels showed that 53% believed STEM education should be implemented in elementary school, while 30% supported STEM implementation in junior high school, 11% supported STEM learning in all grades simultaneously, and 6% advocated for STEM instruction in high school (Tanenbaum, 2016). The work of Gravemeijer, Stephan, Julie, Lin, and Ohtani (2017) emphasized a critical need for mathematics literacy in elementary education to meet the demands of STEM-focused curricula, which students will experience as they progress academically and sit for international assessments (Allen-Lyall, 2018). While there is growing body of literature on the importance of early STEM-based mathematics instruction, only a few studies have researched the effect of early STEM instruction on mathematics gains in elementary school (Doerschuk et al. 2016). STEM disciplines are uniquely interlocked, and the importance of STEM-focused mathematics is observed when students are engaged in activities that promote investigations in engineering, science, and technological principles in which primary grade students thrive (Confrey & Maloney, 2015).

Platas et al. (2016) and the Organization for Economic Cooperation and Development (OECD; 2016a) posited that students who demonstrate early interest and talent in mathematics education are more likely to be self-efficacious and motivated to pursue STEM studies in later years. Gunderson, Park, Maloney, Beilock, and Levine (2018) found that students who are motivated to learn mathematics experience positive learning trajectories from first grade through postsecondary education. Some of the benefits that educational opportunities in STEM disciplines provide include student personal welfare, intellectual growth, and the establishment of a competitive nation on the global playing field (National Academy of Sciences, 2005). While STEM education is a key facet of the U.S. standing in the global economy, a vast majority of the nation's schools teach from a traditional curriculum (National Academy of Sciences2005), and many students who are STEM educated work in non-STEM fields, which increasingly seek STEM knowledge and skills (Grinis, 2017).

A correlation between early STEM instruction and later success in mathematics is evident in previous studies; at the same time, the lack of instruction in mathematics fundamentals leads to low mathematics achievement (Kermani & Aldemir, 2015; McClure et al., 2017; Romar & Matthews, 2015). Children are naturally curious and often exhibit a set of informal mathematical skills before the third grade that researchers have found to be instinctive, broad ranging, and complex, which teachers can tap into with intentional teaching methods such as teaching problem-solving skills in problem-based learning in STEM classes (Daugherty, Carter, Swagerty, & Daughtery, 2016). Educational experts voice interest in early-stage STEM instruction in elementary school based on research that children are naturally inclined towards STEM learning due to their explorative natures and innate interest in mathematics naturally found in their surroundings (Stipek, 2017; Weiland & Yoshikawa, 2013). At the same time, well-designed instruction in mathematics fundamentals supports children's achievement as measured on standardized tests (Darling-Hammond, 2017).

Therefore, the effect of STEM education in the primary school years on mathematics achievement as it is assessed by school districts is an action worthy of study. By studying the achievement of third grade students who were enrolled in STEM education compared to third grade students who were not enrolled in STEM education, I intended to determine if STEM education and its approaches to learning have an effect on student outcomes on the third grade state-mandated mathematics test. This study is important because it can lead to more specific teacher development, improvements to STEM-based elementary curricula, and increased student achievement in mathematics assessments.

Problem Statement

The problem that formed the basis for this study is the lack of information about the effect of early-stage STEM instruction on student state-mandated standardized mathematics achievement test results. Current literature (Nguyen et al., 2016; Schoenfeld, 2016) indicates that educational stakeholders understand the importance of early mathematical instruction to enable children's success in more complex mathematics classes in secondary and postsecondary education. Parents also are concerned about their children's early mathematics learning and achievement because those who are proficient in mathematics tend to advance into high paying technologically based fields (Bailey, Siegler, & Geary, 2014; Fayer et al., 2017; Nguyen et al., 2016). Knowledge about the effect of early STEM instruction on later mathematics success is relevant to educational systems because American students have underperformed in several cycles of international assessments in mathematics and continue to score well below East Asian countries (McDonald, 2016; OECD, 2016b).

There is a lack of research on whether early STEM instruction predicts later mathematics achievement (Nguyen et al., 2016). Previous research findings indicate that only 8% of high school graduates are ready for STEM majors in college, thus affecting the number and quality of STEM talent recruited into STEM careers (Carnevale, Smith, Gulish, & Hanson, 2015; Emeagwali, 2015; Krehbiel & Piper, 2017; Lachapelle et al., 2014; Nguyen et al., 2016; Rothwell, 2013). As a result, more than 80% of manufacturing executives worldwide have expressed concerns about a shortage of STEM talent to meet the exigencies of STEM jobs, given consumer demands of STEM-based products and services (Bryson, Mulhall, Lowe, & Stern, 2018; Holzer, 2017).

STEM instruction in the early grades has received little attention from influential stakeholders (Chiu et al., 2015; Ejiwale, 2013; McClure et al., 2017). This lack of information about the effect of early STEM instruction on children's mathematics achievement is the problem that formed the basis for this study.

Purpose of the Study

The purpose of this quasi-experimental study using retrospective, longitudinal data, and individual growth curve (IGC) models was to determine whether mathematics scores from third grade student state-mandated standardized mathematics test, the dependent variable, differed between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools (non-STEM). In this study, the independent variable of school type was dichotomous because it includes two categories of STEM and non-STEM schools. Since IGC models focus on developmental changes over time (Shek & Ma, 2011), the factor of time (the 6 years between 2012 and 2017) constitutes another independent variable in this study. State-mandated standardized mathematics assessment scores of third grade students formed the dependent variable. The STEM-based schools and the non-STEM schools were in the same school district in a Southwestern state of the United States.

It is important to note for this study that IGC models have two levels of analysis, Level 1 model and Level 2 model, which I used to test two RQs. I analyzed intraindividual and interindividual differences in growth over time based on the results from third grade student standardized mathematics test. The Level 1 model focused on RQ1 and the Level 2 model focused on RQ2. To address the purpose of this study I conducted a longitudinal data analysis using IGC models to answer RQ1 and RQ2 by analyzing results from third grade student state-mandated mathematics standardized tests administered from 2012 to 2017.

Research Questions

Two questions guided this study:

RQ1: What are the individual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$1: There are no statistically significant changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$1: There are statistically significant changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

RQ2: What are the between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$2: There are no statistically significant differences in between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in

STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$2: There are statistically significant differences between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

Theoretical Foundation for the Study

The theoretical foundation that guided this study was Polya's (1957) theory of mathematics problem solving and heuristics. Four key elements of Polya's (1957) state that students must (a) understand what the problem is to determine the best possible method to generate solutions, (b) devise a plan of strategies to solve the problem, (c) execute the plan, and (d) look back on the problem and the outcomes and explore additional paths to the answer. Polya (1957) asserted that the teacher's role is to facilitate learning and find a balance to avoid giving the student insufficient or too much assistance, which fostered student-centered learning and independent thinking. Polya (1957) posited that students usually stop working once they solve the problem and posited that they should work exhaustively to find solutions and new ways to answer mathematical problems. Tanenbaum (2016) and English (2017) noted similar habits in terms of working hard and tenaciously. Polya (1957) discussed trial and error and guesswork as a natural part of working through problems, which Tanenbaum (2016) also mentioned. Past research on teaching mathematics by implementing Polya's problem-solving method in elementary school indicated that students are less likely to abandon the task because of perceived failure and will resort to problem-solving skills in search of solutions (Selmer & Kale, 2013).

Schoenfeld (2013), who has researched Polya's work at length, recalled theoretical analyses he developed on why people succeeded or failed at solving a broad range of mathematics problems. Schoenfeld (2013) discussed four actions that determined a successful problem solver, which stated (a) the student must be knowledgeable about the problem area, (b) students must know possible strategies to find solutions to problems, (c) the student must autonomously regulate his progress and responses to the problem, and (d) the student must practice flexibility when solving math problems. Schoenfeld (2016) found some of Polya's (1957) strategies to be expansive, particularly Step 2, which urged students to decide on a strategy despite the fact that there could be a myriad of strategies from which to select.

Polya's (1957) model provided a framework for this study through which to investigate whether third grade student participation in a STEM-based mathematics course resulted in a statistically significant difference in mathematics test scores compared to those of students enrolled in non-STEM schools. In many cases, traditional methods of teaching math still use rote learning and memorization of facts (Abdullah, Halim, & Zakaria, 2014), which O'Connor, Morsanyi, and McCormack (2018) believed has value when students are developing counting and ordering mathematical skills. Fan and Yu (2017) highlighted the problem-solving process and the engineering design process as effective tenets of a STEM program in the report, *STEM 2026*, which are similar to Polya's (1957) approach to teaching and learning mathematics using problem-solving processes.

According to Tanenbaum (2016), students who repeatedly sought solutions to the challenging problems learned through trial and error and guesswork, but they also used different techniques to pursue answers. Students who were encouraged to follow Polya's (1957) problem-solving process developed persistence, confidence, critical thinking abilities, and metacognitive skills. These attributes are concomitant with quality STEM education. Authentic STEM programs include practical application through hands-on investigation, solution design, collaboration, real world contexts for learning, opportunities to experience failure, opportunities to communicate with other learners, student-centered instruction, and teacher facilitation of student thinking (Tanenbaum, 2016). Polya's ideas are in concordance with these applications and confirm that this theory formed an appropriate foundation for this study.

Nature of the Study

The purpose of this retrospective, longitudinal study using (IGC models was to determine whether mathematics scores from third grade student state-mandated standardized test, the dependent variable, differed between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. I examined state-mandated standardized mathematics test scores of third grade students enrolled in one school district in the Southwestern United States to determine if there was any difference in the growth over a 6-year time period from 2012 to 2017.

The quantitative data I used to answer the RQs were extracted from 2012 to 2017 from a database of scores from a standardized mathematics test that is administered every spring in the district that was the target of this study. These data were publicly available on the analytic portal of the website of the state educational agency responsible for primary and secondary public education and issues related to student testing and accountability. I analyzed the data using a longitudinal, retrospective method utilizing IGC models to determine growth trajectories. STEM-based schools were cluster-sampled and non-STEM schools were stratified sampled from 13,755 third grade students attending 21 STEM schools and 138 non-STEM schools during the academic years of 2012 through 2017.

The intent of achievement testing is to monitor student performance levels based on instruction they experienced in a given subject during the academic school year, as well as to gather data on student academic growth over time (Petscher, Kershaw, Koon, & Foorman, 2014). Given this, the retrospective, longitudinal design in this study was the best fit to examine student performance and monitor student progress. Using a longitudinal, retrospective approach supports observation of repeated measures of the same variables and individuals and has the power to describe the direction of change over time (Caruana, Roman, Hernandez-Sanchez, & Solli, 2015). Employing a longitudinal, retrospective method using IGC models offered the power to reveal any rates of change measured over six time points from third grade student standardized mathematics test outcomes based on STEM or non-STEM instruction. Participant attrition issues such as withdrawal from the study or loss of contact with participant can be problematic when conducting longitudinal studies (Young, Powers, & Bell, 2006); however, given the retrospective nature of this current study, participant attrition was not a concern.

IGC models gained popularity in longitudinal data analysis in educational research due to their strength and generalizability (Willett, Singer, & Martin,1998). They are flexible in nature (Singer & Willett, 2003) and can efficiently model patterns of change over time in student outcomes based on chronicity and timing of the data (Caruana et al., 2015). To analyze data, IGC models must have at least three time periods to assess growth; however, with five or more time periods, such as in this study, which had six time periods, estimation of hypothesized IGCs are possible (Burchinal, Nelson, & Poe, 2006). In addition, using IGC models can facilitate estimation of intraindividual and interindividual achievement growth or lack of growth over time to determine trends in standardized mathematics test scores (Shek & Ma, 2011) of third grade students who experienced STEM or non-STEM instruction.

Generally, an analysis of variance (ANOVA) model is used to analyze changes over time; however, an ANOVA would have been ill-fitted to this study for several reasons. ANOVA requires the use of independent data (Heck, Thomas, & Tabata, 2013), and the data in this study are not considered truly independent due to a higher-level of clustered units, which is time. ANOVA requires the study to have a balanced design (Grilli, Panzera, & Rampichini, 2018), and when performing a longitudinal study, it is common that the nature of the data includes unbalanced data in that there are not an equal number of observations across time periods. In addition, the ANOVA model can only focus on group differences in patterns of growth trajectories, while IGC models can be used to examine change in both group and individual levels (Heck et al, 2013). Because this study had a higher number of waves, meaning the number of time periods (2012 to 2017), IGC models were expected to estimate change parameters with greater accuracy than could an ANOVA (Heck et al., 2013). Lastly, IGC modeling is a more powerful statistical test than ANOVA in that IGC allows detection of individual and group differences that exist within the study, whereas ANOVA is limited to use in finding differences only at the group level (Heck et al., 2013).

Definitions

The following definitions were important to this study:

Individual growth curve modeling (IGC): According to Shek and Ma (2011), IGC modeling is a technique by which a researcher may describe systematic change in individual cases and differences between cases in outcomes over time across distinct measurement waves.

STEM: According to National Academy of Sciences (2017), STEM is an acronym including disciplines of science, technology, engineering, and mathematics, and is used to promote the study of the four disciplines as connected rather than taught in isolation.

STEM pipeline: The National Research Council (National Research Council [NRC], 2013) defined K-12 STEM pipeline as the educational pathway for students ranging from kindergarten to high school who are involved in STEM classes, and/or planning to study a STEM field upon graduation, and extending into the workforce.

Assumptions

This study was based on the assumption that the instructional methods for mathematics followed STEM protocols in the STEM schools and followed traditional protocols in the non-STEM schools. I also assumed that the 20 non-STEM schools randomly selected to represent outcomes for traditional mathematics instruction were representative of the larger population in the target school district of 158 non-STEM schools. I assumed that students enrolled in STEM and non-STEM schools shared similar characteristics and differed from each other only in the instructional model followed at their schools. I further assumed that students enrolled in third grade in both STEM and non-STEM schools were enrolled in the same or similar STEM or non-STEM programs in Grades K through 2, so that third grade test results reflected the cumulative effect of primary grade instruction that was consistently STEM or non-STEM for each child.

Scope and Delimitations

The scope of this study comprised mathematics test scores of third grade students enrolled in either STEM or non-STEM schools within a single public school district in a major city in a Southwestern state. This study was delimited to include existing standardized mathematics achievement test data from third grade students in 18 STEM designated schools and 18 non-STEM schools. Archived mathematics test scores of all third grade students enrolled at the 36 schools from 2012 through 2017 provided data for this study.

Limitations

Several limitations affected the generalizability of findings in this study. One limitation that affected the validity of the results obtained in this study was that students whose mathematics scores were included for data analysis may have experienced different levels of STEM and non-STEM instruction in Grades K through 2, which might have had an effect on mathematics achievement in third grade. Because this was a longitudinal study based on existing data of students from many different classrooms and teachers, it was not possible to confirm the degree to which traditional or STEM instruction was delivered with integrity. It was also possible that parents may have decided to locate their families within the STEM or non-STEM enrollment area of an included school because of personal preference for STEM or traditional education, and this preference may have affected children's learning in unknown ways. In the interest of gathering as large a sample as feasible for this study, there was no attempt to exclude students based on their personal history of STEM education, meaning that children may have experienced from 1 to 3 years of STEM or traditional instruction in their 3 years prior to mathematics testing. These limitations may have affected the validity or generalizability of the results; however, these limitations were offset to some extent by the large size of the data set.

Significance of the Study

The results of this study demonstrated differences in third grade student results on a state-mandated standardized mathematics test based on the type of instruction, STEM or non-STEM, received in early childhood. Consequently, the results of this study inform early educators of the relative value of STEM education and the importance of teacher development programs that focus on STEM instruction. Moreover, the results of this study inform the development of primary grade curriculum regarding STEM-related instruction.

Furthermore, the results of this study are significant because the investigation of research-based programs that facilitate early STEM learning and outcome was warranted (English & King, 2015). Because a bulk of STEM-focused research concentrated on secondary grades, a gap existed in the literature about the benefit of researching STEM education in primary grades (Chiu et al., 2015). With the outcomes of this study, the body of knowledge gained a clearer picture of the effects of STEM education in elementary schools and increased understanding of factors that affect mathematics achievement in the elementary grades.

Summary

STEM education is an integral part of the advancement of a myriad of industries (Mann, Rehill, & Kashefpakdel, 2018), but there is a shortage of talented STEM graduates who might fill positions in those industries (Holzer, 2017). The benefits of STEM education in mathematics achievement and advancement in STEM pursuits are not yet fully understood because of the lack of research on STEM education delivered in the primary grades (Subramanian & Clark, 2016). In this quasi-experimental study using longitudinal, archival data, I determined whether mathematics scores on third grade required assessments administered over a 6-year period from 2012 to 2017 differed between students who participated in STEM-focused education in Grades K to 3 and those who participated in traditional education. In this chapter, I presented the problem of the current lack of understanding of the effect of early STEM education on children’s mathematics achievement. In Chapter 2, I provide a review of literature related to STEM education. In Chapter 3 I describe the quantitative research design and rationale, the methodology, and the plan for data analysis.

Chapter 2: Literature Review

The purpose of this quantitative, retrospective, longitudinal study using IGC models was to determine whether mathematics scores from third grade student state-mandated standardized mathematics test differ between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. Information provided in this chapter further supports how integral this study is to understanding the importance of early mathematics instruction, mathematics as the foundation for working in STEM fields, STEM-focused and non-STEM mathematics education in elementary school, and the connection between early mathematics instruction and later mathematics achievement.

By 2020, unless viable reforms emerge that support early STEM education, only 34% of individuals in the United States will be qualified to fill the 123 million highly-skilled, high paying STEM jobs that will be available in the workforce (Noonan, 2017; Rothwell, 2013, Sithole et al., 2017). Extensive evidence has shown that early-grade mathematical ability in a broad range of skills is indicative of later mathematical achievement (Cerda, Im, & Hughes, 2014; Davies, Janus, Duku, & Gaskin, 2016; Oberle et al., 2014; Rosicka, 2016).

In this chapter, I provide a synopsis of what is known about the effects of STEM education in the primary grades as well as the current gaps in knowledge. I also describe the literature search strategy, explicate Polya’s (1957) problem-solving heuristics as the theoretical foundation of this study, and conclude the chapter with a summary.

Literature Search Strategy

I primarily used the Walden University Online Library to access peer-reviewed journal articles to develop this study. I accessed most of the information through Academic Search Complete and ERIC and Education Source Combined Search. Other databases I utilized were Child Care and Early Education Research Connections, National Academies Press, OECD iLibrary, ProQuest Central, and SAGE Journals. The primary key search terms I used included the categories (a) *early mathematics instruction*, (b) *mathematics achievement*, (c) *problem-solving strategies*, and (d) *STEM education*. Secondary search terms I used in conjunction with the primary key terms were *brain activity patterns, child development, elementary students, later achievement, standardized mathematics test*, and *third grade*. Additional secondary key terms included *elementary school, primary school, early learners, PISA, TIMSS, STEM workforce, traditional education, problem-based mathematics*, *process standards*, and *Polya*.

To validate original resources provided on STEM jobs, STEM attrition, and STEM degrees, I accessed government websites such as the National Institutes of Science, U.S. Department of Labor, and U.S. Department of Education. Pertinent information on the school district from which the STEM and non-STEM schools were targeted originated from its web portal. The majority of the data I used were published within the last 5 years; however, I utilized older reports to provide still-relevant historical perspective.

Theoretical Foundation

Polya's (1981, 1957) theoretical viewpoint on mathematics problem solving provided the theoretical foundation for this study. Polya's (1957) framework for mathematical problem solving and heuristic approaches includes four basic principles of (a) understanding a problem, (b) devising of a plan, (c) implementing the plan, and (d) evaluating the outcome. In his seminal work *How to Solve it* (1957) and *Mathematical Discovery* (1981), Polya defined solving a problem as finding a process that can be applied despite existing barriers and achieving the desired outcome despite initial inability to do so. Polya (1957) prescribed a set of heuristics for each principle for teachers to utilize to guide students through the arduous process of finding solutions to mathematical problems, which Polya believed was recursive and dynamic rather than a set of rigid or linear rules.

Step 1 of Polya's approaches to solving problems is to understand the problem. This first step was designed to prompt students to first read the problem for comprehension and consider prior learning or knowledge about the strand of problem presented. Step 2, devise a plan, taught students how to choose the most appropriate strategy to solve the problem once the conditions of it were established in the first step. Step 3 of Polya's (1957) method, carrying out the plan, required students to implement the best strategy to perform calculations to ascertain answers to the problem. Step 4, looking back, is the final and reflective component of the framework in which Polya (1957) specified students should examine their answers to the mathematical problem they solved and check their results to ensure they responded to all of the conditions of the question and considered the use of alternative techniques in the process (Polya, 1981).

Polya (1957) promoted the use of problem-solving heuristics because he believed that challenging grade-appropriate mathematical problems would boost student curiosity and confidence, develop independent thinking skills, and create an excitement about discovery and inventiveness if they had a proven method upon which to rely. Polya (1957) also believed that his nonlinear approach to solving mathematical problems prepared students with the necessary strategies to handle complex, nonroutine mathematical problems in subsequent grade levels.

In *Mathematical Discovery*, Polya (1981) offered several topics for teachers' use for professional development or as strategies they could implement that would engrain mathematical habits within student thinking, which are:

1. Be interested in your subject: Polya (1981) told teachers that boredom begets boredom; therefore, they should mask any tedium about teaching a familiar topic, as it would diminish student interest in the lesson and math learning.
2. Know your subject: Polya (1981) wrote that interest in teaching a subject was indispensable but not a sufficient condition to teach it, because a lack of knowledge in mathematics meant that students would receive faulty information and methods in the instructional process.
3. Know methods of learning: Polya (1981) explained that providing students with every answer to mathematical problems robbed them of learning how to think as well as depriving them of developing the ability to self-discover.

4. Understand facial expressions of students: Polya (1981) suggested teachers closely watch student physical movements and facial expressions to look for instances where they faced genuine obstacles and needed guidance.
5. Guess: Polya (1981) stated that making reasonable guesses was important in solving problems, particularly in extreme situations when students faced obstacles; conversely, making wild guesses lacked basis and substance.
6. Look Back: Looking back on the work performed to solve the problem was vital, because students verified whether the answers were true or not, which helped students check for errors or determine whether they understood, planned, and executed the problem properly (Polya, 1981).
7. Pattern: Polya (1981) taught that patterns emerged as students solved problems; for example, the process of outlining the problem created patterns as the solver approached the answer. (e.g. multiplication of twos: 2, 4, 6, 8, 10, ___, 14, 16, 18, 20). The student could then make a reasonable guess that the answer in the example would be 12 given that the student observes a pattern of each following answer being two points higher.
8. Analogy: Polya (1981) said that teachers should train students to look for analogous problems by comparing the current problem to similar problems, or looking for analogous approaches to solving the problem if challenges arose in the problem-solving process.
9. Make suggestions: Polya (1981) believed that teachers should be facilitators who did not provide answers for students but allowed them to develop

independent, inductive, and reflective thinking skills by letting them ask questions and provide their own answers to the problem.

These elements have relevance for teachers of children at all ages and grade levels.

While literature on the use of Polya's math problem-solving method in a longitudinal study on third grade student mathematics achievement is nonexistent, limited information is available that mentions Polya's techniques in STEM education and mathematics problem-solving. In addition, however, Griffin and Jitendra (2009) found that techniques from Polya's method are widely implemented in traditional elementary and secondary school mathematics textbooks. As documented throughout this section, Polya's (1957) problem-solving approach was a good fit for this study; therefore, using his heuristics to guide this study was appropriate. In the remaining pages of this literature review, I present ideas relating to the importance of early mathematics instruction, the importance of mathematics education generally, early mathematics instruction in elementary school following a traditional and a STEM model, both generally and in the school district that is the focus of this study, and an overview of assessment of mathematics achievement at the third grade level.

The Importance of Mathematics Instruction

The NRC (2013) reported that mathematical sciences, defined as several disciplines that are not purely mathematical in nature but have mathematical underpinnings, have made major innovative strides in complex applications in computation and digitalization, information technology, and automatization. Mathematical sciences are beneficial to industries that rely on science, technology, and engineering (NRC, 2013). For example, the application of mathematics is integral to many STEM fields, including computer sciences, engineering disciplines, medicine, chemistry, physics, astronomy, defense, and manufacturing. Other nonmathematics areas that rely on mathematical sciences include communications, information processing, and the psychological and social sciences.

There is concern about results from the 2015 administration of the PISA mathematics test that showed mathematics scores for American students were statistically significantly below average for the 35 PISA participating countries worldwide (OECD, 2016a). Alden, Schwartz, and Strauss (2016) and Hausman and Johnston (2014) noted that the 2015 PISA measurement is an indication that America’s global competitiveness might be in decline. The PISA mathematics results from last administration of the test in 2015 showed that the U.S. score of 470 was an 11-point drop from its 2012 average score of nations, which was 481 (Jackson & Kiersz, 2016; OECD, 2013, 2016b). Asian countries including China and Singapore continue to outperform their U.S. counterparts in essential mathematics concepts, skills, and knowledge they should have already learned (OECD, 2016b). Arik and Geho (2017) and McClure et al. (2017) suggested that mathematical training must begin in early education because it is difficult for students to acquire high-level mathematical talent in later educational years. In contrast, Clements and Sarama (2016) found a lack of information about the effect of early-stage STEM instruction on student mathematics achievement.

The Importance of Early Mathematics Instruction

Children exhibit a set of informal mathematical skills before the third grade that researchers have found to be instinctive, broad ranging, complex, and sophisticated (Daugherty et al., 2016). For example, children enjoy using building blocks and can distinguish between different relative sizing and patterns of block shapes and determine sorting, measurement, and order when erecting structures (Stipek, 2017; Weiland & Yoshikawa, 2013). Hands-on learning, including with blocks and shapes, makes children competent in basic geometric skills (Yoshikawa, Weiland, & Brooks-Gunn, 2016; Weiland & Yoshikawa, 2013). Before third grade, students have implicit scientific skills and they grasp basic concepts of physics, amounts and measurement, chemistry, and psychology (Daugherty et al., 2016). Third grade students are flexible thinkers, perform well at collaborating and planning advanced tasks, and can be inventive in creating alternative strategies to solve a mathematical problem (Shoenfeld, 2016). Nunes, Bryant, Evans, and Barros (2015), in a longitudinal study of seven- to nine-year old children's mathematical achievement, found that preschool and primary school children have a strong sense of quantitative relationships, which boosts their ability to make multiple representations in mathematical relationships and problem modeling. Therefore, young children are naturally inclined to learn mathematics and providing them with opportunities for formal instruction in mathematics throughout the early years makes sense (Hefty, 2015).

According to Watts, Duncan, Siegler, and Davis-Kean (2014), mathematics is an incremental discipline, with understanding of advanced concepts dependent on basic understanding established earlier. Mathematics education has depended on the idea that students learn new information presented in a particular school year, and also build on their prior mathematics knowledge, since students usually cannot solve advanced mathematical problems without having learned basic mathematics processes in earlier grades (Watts et al., 2014). According to Harris, Petersen, and Wulsin (2016), exposing young children early on to age appropriate mathematical concepts related to numbers and emergent counting, sorting, patterns and shapes, and measurement, supports higher mathematics skill development, application of mathematics skills to solve problems later on, and confidence in solving mathematics problems. Watts et al. (2014) reported that early grade mathematics ability can predict mathematics achievement in adolescence. Nguyen et al. (2017) found that preschool students who mastered counting skills were advanced mathematics students throughout elementary school. Therefore, mathematical training must begin in early education so that students might acquire high-level mathematics achievement in later educational years (Arik & Geho, 2017; McClure et al., 2017).

Vertically Aligned Performance Standards from P-16

Vertical alignment ensures that students state-learn required knowledge and skills as they progress from one grade to the other (Moore et al., 2014). The state of focus in this study is one of 44 states in the union with state-mandated vertically aligned curriculum standards (Schoenfeld, 2016). Vertically aligning standards from grades P-16 is one of three initiatives that involves the state’s educational association and workforce commission in the effort to connect primary, secondary and post-secondary educational systems to provide students with a solid mathematics foundation from early years into the workforce (Daily, 2014). P-16 mathematics standards are aligned sequentially with the curriculum, as well as to the state's assessment instrument (Daily, 2014).

The educational association in the state that is the focus of this study designed the mathematics curriculum in a comprehensive system that vertically aligned the curriculum and performance standards starting with college and career readiness standards in high school, down through elementary grades, and then projecting forward throughout college in order that students broaden their skills in each subsequent grade (Polly & Orrill, 2012). The reverse design in the vertical alignment was intended to ensure that students are prepared to function successfully throughout postsecondary education and compete on the global stage.

The school district that is the focus of this study is within the target state's energy and STEM corridor, which is one of the largest in the nation, and prides itself in preparing STEM-talented students. One of the reasons for the school district's vertical alignment map was to establish an approach to better deliver standards that more successfully connect elementary student achievement with continued success throughout high school in its effort to decrease the high school dropout rate. The school district's high school dropout rate, according to the target state's Academic Performance Report (APR) from the 2016-2017 school year, was 16.1%, which concerns state officials.

Every student enrolled in grades 3-8 in the target state is required to pass the mathematics high-stakes standardized test in order to achieve promotion to subsequent grades. Scores from the 2017 mathematics standardized test show that 73% of third grade students and 67% of seventh-grade students scored in the Approaches Grade Level (AGL) range, while 79% of eighth-grade students and 73% of high school Algebra I students managed AGL expectations. The AGL range mean that these students met the minimum standards and will be promoted, but are likely to require targeted skill building intervention in order to be successful in mathematics in later grades.

The APR of the targeted school district is 46% which means that the total population of students in the district who took the standardized mathematics test scored in the Meets Grade Level (MGL) range. Students in the MGL range demonstrated that they understood the subject matter and ready for postsecondary studies, but some may receive minimal, targeted interventions, as the state believes that there is room for improvement). The 2017 mathematics results concern educational stakeholders because extensive efforts were invested into establishing a comprehensive vertical alignment system with fewer, more rigorous performance standards. Nonetheless, statistical results from the targeted school district show that a high percentage of students lack understanding of specific grade level content knowledge and skills they should have been taught starting in the lower grades.

While performance standards provide a framework that defines what knowledge and skills teachers must reliably teach, the most effective instructional strategies that successfully connect student achievement and academic development with the standards are separately determined by each school district (Chang & Silalahi, 2017). Determining instructional approaches that meet the needs of the diversity of students enrolled in a large school district is a major undertaking (Dolan & Collins, 2015; Schanzenbach, 2014). With that said the instructional tools and techniques that the district implements to meet student learning needs based on the standards do make a difference in terms of individual student performance and improving student achievement (Booth et al., 2017; Koedinger, Booth, & Klahr, 2013). Instructional methods employed by the district that is the focus of this study will be discussed next.

Early Mathematics Instruction with Traditional Focus

Traditional education has longstanding practices dating from the early years of the one-room schoolhouse in the late 1800s where teachers were the central figures of knowledge, which they directly dispensed to students who were considered passive learners (Dewey, 1915) and receptacles of transmission of knowledge, which Freire (1996) characterized as the banking model of learning. The National Council of Teachers of Mathematics (National Council of Teachers of Mathematics [NCTM], 2000) reformed its prescribed standards for K-12 mathematics education in 2000 to indicate a shift towards a more active learning, constructivist method of learning (Moody & DuCloux, 2015; NCTM, 1995, 2000, 2006). The traditional focus of learning early mathematics centers on repeated and rigorous arithmetical computations and memorization of mathematical proofs to develop basic problem solving skill. Prototypically, traditional teaching methods consist of students exhaustively practicing learned algorithms, which promote memorization of facts (Boaler, 2015). This motorized model of learning that Ono (1966) experimented with showed that habit-strength acquisition practice times within a controlled situation is Pavlovian in nature, as per Hull (1943), and akin to the modern “drill and kill” learning conditions practiced in the today’s classrooms.

The NCTM (2000) also altered its principles and standards to promote newer methods of communicating mathematically because it argued that the demand of technological advancements in most industries, and the aftereffects on student development, performance on standardized and international assessments, and their ability to live and work successfully in real world occupations in adulthood was eminent (NCTM, 2010; OECD, 2017). Furthermore, NCTM determined that traditional teaching of early mathematics was not sufficient to push students and teachers out of their comfort zones, thereby, questioning the effectiveness of its strategies (NCTM, 2006; Nguyen et al., 2016). Despite the change, traditionalists still believe that modernized standards and instructional strategies supported by NCTM undermine traditional teaching cultures, and teacher experience (Nguyen et al., 2016). In addition, opponents of reform mathematics have not easily relinquished the initiator and controller factor of teaching mathematics in exchange for a more facilitative approach with students (Fullan, Langworthy, & Barber 2014; Lameras & Moumoutzis, 2015). Traditional approaches to teaching early mathematics is controversial to certain stakeholders and policymakers because the methods are teacher-focused, and some believe that the "sage on a stage" approach robs students of deeper learning activities that shape the ability to communicate mathematically, observe relationships in patterns, and enhance projective, collaborative and divergent thinking (Lithner, 2017).

Compliance oriented students become used to nonassociative learning which is linked to the Pavlovian theory of classical conditioning, in that students develop behaviors to particular stimuli which are formed in response to repeated events overtime, for example, in a classroom (Anderman, 2010; Bray & Tangney, 2017; Pavlov, 1927; Rescorla, 1988; Skinner, 1953, 1957, 1969). In traditionally instructed classrooms, student collaboration and solving mathematical problems that have been designed to reflect real world contexts are nearly non-existent. Instead, students focus mainly on content-oriented processes (Castronova, 2002) such as mastering fixed formulas and basic algorithms (Bray & Tangney, 2017; Lameras & Moumoutzis, 2015; Maab & Artigue, 2013). Memorizing quantitative procedures to later master assessments (Haridza & Irving, 2017) reduces the need for student feedback, interpretation, or discovery (Castronova, 2002). Mathematics accounts for 60% of the curriculum in Chinese schools, and while China is often criticized for implementing rote learning and memorization as the sole means of teaching early mathematics, it continues to flourish as a top performer on international exams (Zhao, 2014).

In traditional methods of teaching early algebraic mathematics as found in pre-packaged lessons in textbooks, students are normally provided an example problem, given step-by-step explanations for each variable in the problem, and armed with memorization of formulas and algorithms, students are provided worksheets containing problems (Corlu, 2013). Traditional methods of teaching early mathematics are still practiced today in many parts of the U.S. and other countries, including parts of East Asia and Europe, many of which, in comparison to the U.S., perform better on the PISA mathematics test (OECD, 2016c). Proponents of traditional education insist that its conventional approaches to learning mathematics promote student achievement and compete with reform mathematics techniques to teaching and learning early mathematics, and are equally effective in meeting the demands of rapidly advancing and emerging technologies (Mbodila & Muhandji, 2012).

Traditional Mathematics Instruction in the School District of Focus

In the school district of focus in this study, traditional mathematics is delivered by general education teachers through a published curriculum using the textbook, *GoMath!* Students receive instruction via traditional textbook methods; however, the curriculum has a digital component that provides opportunities for interactive activities online through *K-6 Think Central*. The district adheres to mathematics standards as outlined by the educational association of authority in the state, and requires general and special education teachers to create relevant learning experiences based on student backgrounds at home, work, recreation, and leisurely interests. The district's curriculum department trains general and special education teachers to use effective math instructional strategies and systematic assessments that gauge student achievement. District policy established 90 minutes a day for third grade students to receive mathematics instruction and practice using the *Go Math!* platform, and buoys students who require supplementary help or intervention with additional time in math tutorials.

The goals that the district established for each traditionally taught school is to build a foundation of basic mathematics understanding in each of the focal areas including numbers and numerical relationships, arithmetic computation and algebra, geometry, measurement, processes of data analysis and consumer math. In order to solve problems in each focal area, third grade students are expected to (a) manipulate numbers up to 6 digits and solve sums and differences using graphs, number lines, and algorithms;

(b) demonstrate understanding of what it means to multiply or divide whole numbers; and (c) be able to manipulate fractional parts of wholes and also to name and sort geometric figures and solids.

Early Mathematics Instruction with STEM Focus

According to Kelley and Knowles (2016), teachers can deliver early mathematics instruction from a STEM-based perspective, using pedagogical practices that involve investigative inquiry (Kelley & Knowles, 2016). STEM-focused mathematics can be achieved through the process of investigating mathematics concepts that are linked to engineering projects that nurture student abilities to develop mathematical skills (Kelley & Knowles, 2016). In addition, teachers can provide quality mathematics instruction by involving students in a variety of learning methods that include hands-on activities and active participation, such as in project- and problem-based learning (Denson, Austin-Stallworth, Hailey, & Householder, 2015) and by permitting collaborative work (Kelley & Knowles, 2016). In addition, teachers can present mathematics problems with ill-structured themes, similar to the complexity of real-world challenges that cultivate student sense-making capacities and promote mathematical inquiry (Fielding-Wells, Dole, & Makar, 2014). When students are involved in solving problems, it forces them to develop supportive argumentation in explaining and defending their ideas and so they learn to negotiate collaboration, and to enhance their skills in comparison, reasoning, and analysis as they apply mathematical principles (Fielding-Wells et al., 2014; Sullivan, Clarke, & Clarke, 2009; Zembal-Saul, McNeill, Hershberger, 2013).

Inquiry-based learning, involving student active participation in the learning process, can be designed through a variety of activities that are engaging, relevant, involve teams, and are based on real life events and complex situations (Freeman et al., 2014; Rissanen, 2014). Incorporating opportunities for primary students to develop mathematical skills through technology and engineering projects establishes a culture of active learning and inquiry in the classroom that helps students develop mathematical thinking skills (Honey et al., 2014; Kelley & Knowles, 2016; Kennedy and Odell, 2014; NRC, 2014). According to Fielding-Wells et al. (2014), a case study with fourth-grade students showed that student mathematics abilities in number operations, fractions, ratios, recognizing patterns, measurement, and comparative reasoning increased through work on an inquiry-based project. Operationalizing mathematics education through project- and problem-based learning supports student math sensibilities when purposeful mathematical concepts and problems are interwoven into the project and contextualized to real life (Fielding-Wells, 2014; Sullivan et al. 2009).

Mathematics process standards provide a framework by which teachers help students enrolled in inquiry-based learning to acquire the cognitive and problem-solving skills needed to solve a range of mathematics problems, such as required on the standardized mathematics test. For example, a study of robotics education with early childhood learners determined the effects on computational thinking involved in programming a robot to perform a dance (Bers, Flannery, Kazakoff, & Sullivan, 2014). The robotics curriculum included major aspects of engineering and computer science principles, two domains that involve mathematical thinking, and started with a lesson on the engineering design process, that provided a framework for planning, testing, and making improvements throughout the project (Bers et al., 2014). Additional activities included debugging the robot, following programming instructions, and controlling flow attributes. The process of programming the robot to perform a dance enhanced student ability to apply suitable solutions to solve problems. They were able to design a plan, and troubleshoot unexpected problems by debugging. The tasks of programming the robot's movement was accomplished with the use of symbols, which developed student symbolic language and mathematical communication and representations. Students also strengthened their number sense and estimation ability through control flow tasks, and used procedural thinking skills to follow sequencing instructions, which rely on order of operations knowledge (Bers et al., 2014). Through rich, hands-on activities such as the robot project, STEM-based education provides a mechanism by which even young children acquire skill in mathematical thinking and computation.

STEM Approach to Mathematics Instruction in the State of Focus

The content standards, which are the essential skills and knowledge that students should possess in order to solve mathematical routine and non-routine problems, are the same for the student populations across the state, regardless of school type (Opfer, Kaufman, & Thompson, 2016). However, content standards are delivered differently in STEM schools than they are in the traditional, non-STEM schools (Opfer et al., 2016). Process standards require teachers to train students to identify, understand, apply, and create ways to find solutions to complex problems. According to a district mathematics specialist, STEM programs incorporate real world contexts that students experience through field trips, which promote collaborative and investigative learning, which are a mainstay component of STEM education and mathematical achievement (McDonald, 2016). The district created its STEM program to promote mathematical literacy through an alternate model of education that was in addition to its general transmissive educational models used in most of its schools, which involve learning procedural facts and algorithms through rote learning and memorization methods. Project and problem-based learning opportunities provide a vehicle by which to promote mathematical thinking through hands-on projects that are challenging, fun, and interesting (Fielding-Wells et al., 2014; McDonald, 2016).

Standardized Mathematics Assessment of Grade 3 Standard Categories

The primary focal areas of third grade mathematics essential knowledge and skills (EKS) include basic arithmetic operations, including manipulation of place value, and fractions. The three focal areas of the third grade essential knowledge and skills are supported by math problems in number and numerical operations, measurement, geometry, elementary algebraic understanding, and problem solving through processes of analysis. Non-routine problems based on place value on the third grade state-mandated standardized mathematics test; for instance, will prompt students to begin implementing specific approaches to problem questioning system in order to spur active and critical thinking, which would help them to avoid mistakes and faulty assumptions about the terms and conditions of the given problem. Making annotations throughout the problem solving process around formulas or drawn tables or diagrams Polya (1957) helps students develop written and oral communication skills, and enhances mathematics vocabulary and computational capabilities in the third grade Powell et al. (2017), which are required in the grade level essential knowledge and skills.

In 1989, the NCTM released *Curriculum and Evaluation Standard, for School Mathematics*, which set uniform standards for students to learn math content objectives, and by which it expected math programs to perform. By 1995, NCTM decided to enhance its standards that included the higher order functions of problem solving strategies to solve math problems, and process standards were implemented (Davis, Choppin, Drake, Roth McDuffie, & Carson, 2018). The process standards were more demanding criteria, which provided a structure for the teaching and learning of mathematics with an emphasis on development of reasoning and skills of analysis (Meltzer, 2018). NCTMs process standards include (a) the ability to solve problems, (b) principles of reasoning and proof, (c) the ability to express mathematics ideas verbally, (d) the ability to make connections between ideas, and (e) the ability to represent mathematics ideas, which the state of focus in this study incorporated into its Grades 3-8 essential knowledge and skills (NCTM, 2014). The NCTM developed five content standards based on core functions of mathematics that students are expected to learn include (a) numbers and operations, (b) algebra, (c) geometry, (d) measurement, and (e) data analysis and probability (NCTM, 2014). The NCTM believed that quality instruction based on problem solving strategies created a solid foundation for all students to learn math in a world increasingly driven by quantitative decisions, and developed a set of process standards it deemed would prepare students for the 21st century (Leong & Janjaruporn, 2015).

Even though the NCTM standards were not nationally mandated, several states, including the target state designed grade level mathematics standards on the association's standards, as a result, the essential knowledge and skills place emphasis on the process standards since they are cognitive function standards based on Polya's four-step framework, and engages students in higher order thinking to solve complex mathematics problems (NRC, 2011). The state educational association incorporated process standards into its third grade essential knowledge and skills, which requires students to use problem-solving strategies to solve math problems based on each mathematics focal area on the mathematics standardized exam. According to Leong and Janjaruporn (2015), the NCTM based its industry-wide mathematics process standards on problem-solving strategies found in Polya's (1957) book, *How to Solve it*, and the essential knowledge and skills listing contains mathematics process standards that are based on the NCTM's Polya-based problem solving strategies. The school district implements the essential knowledge and skills in the effort to provide methodical thinking and problem-solving strategies to students to solve complex mathematics problems that might have a positive effect on the standardized test scores.

The overarching goal of the state's third grade mathematics process standards is to teach students how to apply and use mathematics in solving mathematics problems. There are seven process standards that the district of focus in this study has been using since 1997, which are illustrated in Table 1.

Table 1.

Mathematical Process Standards

		Grade 3 mathematics standards in the focus state
Application of mathematics	A	The student uses mathematics in solving problems that arise in everyday situations
	B	The student uses problem-solving strategies to analyze information and formulate a plan to address a problem, to solve the problem justifiably, and to evaluate their own thinking and the suitability of the solution.
	C	The student manipulates objects, creates drawings or notes, or employs technology, and applies heuristics, to solve problems.
Understanding of mathematics	D	The student discusses mathematical concepts, solution paths, and alternatives, using words, manipulatives, and two-dimensional representations.
	E	The student records mathematical thinking coherently and can share their thinking with others.
	F	The student analyzes conceptual relationships to make connections between mathematical ideas.

The process standards aid students in operationalizing the next group of expectations, knowledge and skills statements, which are the five different mathematical areas that students are taught and are assessed on as shown in Table 2.

Table 2.

Grade 3 Reporting Categories and Benchmarks

Number and Operations	The student uses mathematics in representing and comparing whole numbers and demonstrates understanding of place value
	The student uses mathematics to demonstrate understanding of fractions
	The student uses mathematics to solve problems using computational algorithms accurately and efficiently.
Algebraic Reasoning	The student uses mathematics to observe and describe conceptual relationships and patterns.
Geometry & measurement	The student uses mathematics to describe geometric figures of two-dimensional and their characteristics
	The student uses mathematics to solve measurement problems involving customary and metric units and tools.
Data Analysis	The student uses mathematics to create, organize, and interpret data needed to solve problems, and to present solutions.
Personal Financial Literacy	The student applies mathematics to problems of getting and spending money and managing finances, and demonstrates understanding of basic financial concepts.

The mathematics standardized test has 32 questions, which are built around the five reporting categories. There are eight questions on number and operations, 13 questions on computations and algebraic relationships, seven questions on geometry and measurement, and four questions on data analysis and personal financial literacy. The composition of this assessment was important in analyzing the results of this study.

Summary and Conclusions

In this chapter, I reviewed literature relevant to my study of the effect of mathematics pedagogy on third grade student mathematics achievement. I presented a rationale for this study, given the importance of mathematics ability for the American workforce, and also presented information about mathematics curriculum and the differences between traditional mathematics pedagogy and mathematics instruction in a STEM focus. In Chapter 3, I will present the method by which I conducted this study.

Chapter 3: Research Method

The purpose of this quasi-experimental study using retrospective, longitudinal data and IGC model analysis was to determine whether mathematics scores from third grade student state-mandated standardized mathematics differ between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools (non-STEM). For this current study, I chose a longitudinal, retrospective design using IGC modeling to statistically explain any interindividual and intraindividual changes of mathematics test scores of third grade students who were enrolled in STEM-based schools and students who were not over six time periods between 2012 and 2017.

In Chapter 3 I explain the rationale for the study design and the process by which I accessed the archival data and analyzed them. I discuss threats to internal, external, and construct validity, and the potential for any moderating or mediating situations that may influence the study outcomes. Lastly, I elaborate on procedures to avoid any ethical issues relating to this study.

Research Design and Rationale

Study Variables

The independent variable in this study was the use of STEM or non-STEM mathematics pedagogy. The dependent variable was student mathematics scores on the state-mandated standardized test administered at the end of the third grade year.

In this longitudinal study I used IGC analyses to measure changes over time at both the aggregate and the individual perspectives. There are two levels in IGC models. The Level 1 model is used to test for interindividual changes over time and precludes predictor variables (Shek & Ma, 2011). For this study, Level 1 was focused on each individual school mean mathematics scores to describe changes in average scores over time. For instance, IGC analyses will capture how a school has performed against itself over time (Shek & Ma, 2011). Level 2 will capture whether the rates of change vary across individual schools in a systematic way. In this study, Level 2 captured whether the rate of change varied across individual schools in a systematic way (Shek & Ma, 2011). Essentially, the model shows how schools perform against each other over time. By analyzing both the interindividual differences and the intraindividual changes over time, I hoped to determine if there are statistically significant differences in children's mathematics scores over time, dependent upon STEM and non-STEM pedagogy. IGC analyses do not require balanced data across different waves of data, such as unequal sample sizes, missing data, or inconsistent time intervals (Shek & Ma, 2011).

The school district, as of the 2017 testing season, included 159 elementary schools enrolling a total of 13,755 third grade students. Test data are publicly available on the state-run website of the educational agency and were used to determine if there are any differences in the mathematics scores of third grade students enrolled in STEM and non-STEM educational methods.

Research Design and Research Questions

I used a nonexperimental design to answer the RQs. Nonexperimental approaches do not allow the researcher to actively manipulate the independent variables, and participants cannot be randomly assigned to groups (Cook & Cook, 2008; Creswell, 2013). Nonexperimental design in longitudinal research is a good fit because retrospective research involves analyzing old data and comparing it to present data from the cases in the data set (Johnson, Figueroa-Colon, Huang, Dwyer, & Goran, 2001). Longitudinal research is used to analyze educational data such as test scores from the same subjects over two or more waves and can predict event occurrences over different waves (Henson & Hinerman, 2013). I studied the rate of change between mathematics test scores using inferential statistics from longitudinal data analysis using IGC models. IGC is an advanced technique that I used to examine changes in student mathematics scores across time. IGC techniques modeled systematic changes within STEM schools and within non-STEM schools, as well as between-school mathematics score differences across a 6-year period. The term *individual growth curve* is frequently used to examine aggregates of individual curves instead of separate analysis of each IGC (Shek & Ma, 2011). For example, I examined the aggregate score of all STEM schools over each year, and the aggregate of all non-STEM schools over each year, rather than the aggregate of each school individually over each year. In this study, I tested whether pedagogy for STEM and non-STEM education is predictive of student state-mandated mathematics test scores and determine the trajectory of student achievement in math scores across time.

Time and Resource Constraints

IGC models do not have assumption constraints such as the ANOVA models. I did not experience any time and resource constraints because the data including third grade student state-mandated mathematics test scores for the 6-year period covering 2012 to 2017 was publicly accessible through the state-run website of the educational agency analytic portal, the Assessment Management System (AMS). There were no costs involved to obtain and download the data through the AMS. The time involved to analyze the data once it was downloaded from AMS was consuming. I entered data from AMS into Excel and imported into SPSS. To prepare the data for SPSS analysis, I converted information from wide format to long format where each row represented a school and the wave periods were represented in the proper columns.

I chose a longitudinal study using IGC analyses to investigate differences between STEM and non-STEM standardized mathematics scores to advance knowledge in the discipline because there was no 6-year study of third grade student performance on the state-mandated standardized mathematics test in the largest school district in the area based on whether they are STEM educated or non-STEM educated.

Methodology

Target Population

The target population encompasses predetermined elements to be observed in the study (Daniel, 2012). The observation in this study was the standardized mathematics test scores of third grade students in a public school district in a Southwestern state in the United States. Specifically of interest in the present study were the standardized mathematics test scores of third grade students who were enrolled in one public school district's STEM and non-STEM elementary schools. Third grade student test scores were examined longitudinally across six different time periods from 2012 to 2017. A district-wide relevant population structure of 13,755 third grade student standardized mathematics test scores was divided into two sample groups, STEM and non-STEM. There were 159 total elementary schools including 138 non-STEM campuses and 21

STEM campuses. The protocol I used to select the school district and its STEM and non-STEM schools is described in the following section.

Sampling and Sampling Procedures

I used a one-stage cluster sampling technique to identify eligible participant data. A sample is a subset of a population (Creswell, 2013), and one-stage cluster sampling is used when clusters of all participants that represent the population are identified and included in the sample. IGC models handle clustered data as they measure patterns of mean-level changes over time as in longitudinal studies. Clustered sampling can be drawn in two or more stages, which is common to survey sampling, but in one-stage studies, random clustered samples of schools, gender, or achievement scores are common (see Hedges & Rhoads, 2010).

Sample clustering procedure and sampling frame. There were 21 STEM elementary schools and 138 non-STEM schools in the district that was the focus of this study. The sampling frame for the study included 18 STEM schools because three of the 21 STEM schools did not meet eligibility requirements as described in the next section. While the STEM schools were cluster-sampled, the non-STEM schools were stratified sampled in order to select 18 schools from the pool of 138 to meet the assumptions of homogeneity. Each cluster in the non-STEM school group was assigned a number from one to 138 because each school must have an identification code and cannot be assigned to more than one cluster. I used the Longpower package to conduct a power analysis in order to determine the effective research sample size because the Longpower package is designed to compute linear models of sample longitudinal designs (see Donohue, Garnst, Edland, & Donohue, 2013).

Inclusion and exclusion criteria. I identified eligible schools for this study from information from the state-run website of the educational agency responsible for student testing and accountability in primary education. I also obtained information on school progress and student achievement from the website of the school district in which this study was conducted. The school district had 256 total schools, including 138 elementary non-STEM schools and 21 elementary STEM schools. Three elementary schools in the STEM school population were purged from the sample. Two schools were eliminated because they were mixed education facilities, simultaneously housing elementary and middle school students, which violated the assumption of homogeneity. Furthermore, elementary school students learning in the same facility with middle school students could potentially present biological, social, or cognitive variables that could have undue influence on student academic performance, especially for females (Simmons, 2017). While many schools house elementary and middle grades together, Simmons (2017) reported that early adolescence (12- to 14-years-old) can be tumultuous; Dockrell et al. (2017) suggested this could negatively affect younger student academic achievement. The second school was eliminated from the study sample because it also violated the assumption of homogeneity, as it was the only non-Title I elementary school in the district, meaning its student population comprised a higher socioeconomic status than the other schools. The exclusions and delimitations made here helped to make the results more accurate and to create a sampling cluster that was as error free as possible (see Caruana et al., 2015).

Power analysis. Power analysis determines the best sample size of the study, which helps to conclude statistical significance (Heck et al., 2013). I used Longpower package R software, version 3.4.2, to calculate the power analysis and determine the optimal sample size.

Data Collection

The data I needed to conduct this study were publicly available. The targeted school district was the most appropriate school district because it had a magnet program with distinction schools, such as STEM, non-STEM, Montessori, International Baccalaureate, fine arts, and other school type designations. In addition, the school district is large, registering 256 schools with over 200,000 students, which ensured that this study had ample sample size. Because this current study focused on outcomes of student performance hinging on STEM and non-STEM instruction, this study featured STEM and non-STEM schools in the district. Information about the target school district was located on its home page, which was publicly accessible. Because a goodly amount of current information on each school was provided on the school district’s website, and statistical data on all schools, mean averages, gender, and other demographics were available on the analytic portal of the state-run website of the educational agency, overall recruitment processes that involved locating, enlisting, and selecting participants were not necessary.

I placed a telephone call to the administrative offices of the educational agency and explained my intent to use mathematics test data from Grade 3 for a longitudinal research study. The operator transferred me to the Department of Assessment and Accountability, which connected me to the voicemail of the director of the Division of Performance Reporting. The division informed me that the Data Interaction for Student Assessments through the state-run analytic portal page was not password protected, and contained only standardized test data of overall school district performance was accessible to any person in the public. Individual student names and other biographical data were sealed. I was advised of the *Analytic Portal Help Guide* that is the downloadable user manual, which helps any person to navigate the website for access to analytical data from any school district in the state. Standardized assessment data are archived starting from the 2012 testing cycle.

In the Analytic Portal, the testing program, grade level, years tested, subject tested, individual organization, state, and individual schools can be selected. I selected the standardized testing program for grades 3 to 8, and then clicked the Grade 3, which is the grade level for my study. I selected the spring testing seasons for the years 2012 through 2017. I selected mathematics as the subject and typed in the name of the school district. There was an option to get the full report from the school district's website on particular groups of students based on specific variables of interest. The AMS system produces a full report of group summary by performance levels including the name of the school district, its identification number, the years tested, grade level, number of students tested, average scale score, and the performance levels of each year by satisfactory, advanced, unsatisfactory, or did not meet, approaches, meets, and masters. The data can be filtered by disaggregating into the following subgroups: Gender, ethnicity, economically disadvantaged, Title I, Part A, migrant, Limited English Proficiency, bilingual, English as a second language, special education, gifted and talented, at risk, and military connected. All participants included in the study may not have documented standardized math test scores from 2012 to 2017 with the educational agency. Since standardized mathematics test records are accessible on the public portal, informed consent for each subject in the study was not required.

Instrumentation

The instrument that the state of focus in this study uses to measure third grade student performance in mathematics is the newest designed assessment instrument to gauge academic readiness, established in 2012 as a comprehensive accountability system to increase the rigor of assessing student knowledge in mathematics and other core subjects, and to improve the educational system. All students in state are required to take the standardized test, which assesses student knowledge on content standards as found in the essential knowledge and skills specifically in order to prepare students for postsecondary readiness. The state education association devised its accountability instrument based on strict standards for authentic assessment and accountability predicated on a number of state laws related to standardized testing, and assessment tools and instruments. The state education association collaborated with Pearson Education to develop the mathematics instrument, which is directly aligned with curriculum essential knowledge and skills. Since the state education association owns the instrument, and all public school districts in the state must administer the standardized tests and report disaggregated results, no requirement was necessary to seek permission to use the instrument.

Standardized Testing Program Instruments. The standardized testing program includes annual assessments for mathematics and reading in grades 3 through 8, writing in grades 4 and 7, science at grades 5 and 8, and social studies for grade 8. The testing program Alternate 2 instrument assesses the same levels and subjects, but is an accommodated format for students in grades 3 through 8 who receive special education services. The Online Testing Platform is also an accommodated version that assesses students in grades 3 to 8 who receive special education services, or have cognitive disabilities, in all standardized tested subjects. The standardized Spanish assessment is available for students in grades 3 to 5 who participate in bilingual education programs, while becoming proficient in the English language. The third grade mathematics standardized assessment, which was the focus of this study, is delivered in mostly paper-and-pencil format and includes 32 test items based on the mathematics categories including (a) numbers and numerical relationships, (b) arithmetic computation and basic algebraic ideas, (c) geometry and measurement, and (d) analysis of data and application of mathematics to everyday problems of finance.

The 32 test questions are based on the relationship between the mathematics content or readiness standards and process or supporting standards that students must understand in order to solve the more rigorous and non-routine mathematics problems. Category 1, numerical representations and relationships has eight questions on the standardized instrument, Category 2, computations and algebraic relationships, has 13 questions, Category 3, geometry and measurement, has seven questions, and Category 4, data analysis and personal finance literacy, has four questions. There are a total of 13 readiness standards and 31 supporting standards, which the test assesses. There were 29 multiple-choice questions and three questions requiring students to use a grid to record answers.

Standardized Test Performance Levels. The federal accountability law, the NCLB Act, required the state education association to establish at least three achievement levels as a way to determine satisfactory achievement, and establish performance indicators when reporting and categorizing student levels of performance on the standardized test. The state education association worked with higher education coordinating board of the target state to assemble a Performance Descriptor Advisory Committee (PDAC), consisting of a diverse panel of seasoned educators from public education and higher education, as well as professionals from education advocacy groups to establish three performance levels, define them, and create guiding policies for each level. PDAC was careful to create labels that represented each student's performance level in the appropriate corresponding category, establish labels that represented each performance level, and ensuring that the performance labels focused on guiding policies rather than on student performance. The state education association and the higher education board provided the PDAC the research information and data based on empirical evidence to facilitate the validation of the standardized instrument. The committee reached a consensus after a two-day brainstorming and planning meeting, and recommended three levels of academic performance: 1) Level III: Advanced; 2) Level II: Satisfactory; and Level I: Unsatisfactory.

Third grade students who have developed reasoning and evaluative skills and can apply mathematics process skills to solve non-routine mathematics problems that involve adding and subtracting whole numbers, linear measurement, and observing relationships between mathematics operations are performing on Level III. By the spring third grade math standardized test administration, students who have learned to describe geometric figures and fractional equations using technical terms, solve basic arithmetic problems, identify patterns in related number pairs such as 10 + ____ = 20. Mathematics problems that involve measurement are achieving at Level II. Level I performance on the third grade math standardized test indicates that students have not developed mathematics proficiencies beyond recognizing fractional problems, or up to three-dimensional geometric shapes, symmetric lines, and congruent shapes, as well as using math models counting U.S. currency, or to identify multiplication or division patterns in mathematics sentences.

The standardized third grade mathematics instrument has 32 question items that are linked to Grade 3 essential knowledge and skills, which were redesigned to promote mathematics fluency on a level that requires a list of complex thinking skills that students should have developed before and throughout the third grade to solve the problems successfully. The state education association provided a list of complex cognitive skills, which include analyzing the problem, implementing problem-solving skills to solve non-routine math problems, developed conceptual knowledge, procedural fluency, applying strategic and adaptive reasoning, communicating and justifying responses, and persistence. Of the total of 32 questions on the mathematics instrument, 24 questions must be answered correctly to pass, and 28 questions must be answered correctly to achieve mastery level on the test. Students have four hours to complete the third grade mathematics test.

Reliability and Validity. Reliability in quantitative research refers to the accuracy of the consistency and dependability of the measurement instrument and whether testing and retesting will yield the same results every time that it is administered in the same setting with the same participants at each interval (Sullivan, 2011; Creswell, 2013). Typically, internal consistency reliability is analyzed in large-scale educational assessments, such as the standardized mathematics test to determine how well test questions link to the essential knowledge and skills and measure what they are intended to measure. The state education association discussed the importance of the design of the third grade mathematics instrument, which was intended to adequately measure the essential knowledge and skills at the highest achievement level. Educators split the standards into readiness and supporting standards to ensure a clear connection between what the essential knowledge and skills required students to know. However, content linking was not sufficient to ensure validity; therefore, the state education association ensured that test items on the mathematics instrument were aligned with the higher cognitive complexity, and the mathematics test included open-ended items to assess student ability to think and solve problems independently.

The readiness standards are prominent on the mathematics assessment because they are essential knowledge and skills designed to develop student knowledge on current and subsequent grade levels. The content standards measure what students need to know for promotion to the next grade level. While instruction is predicated on the supporting, or content standards, not all of the supporting standards were included in the mathematics assessment. The mathematics assessment tested grade-level content standards and not an accumulation of essential knowledge and skills standards from previous grades. Item analysis is performed annually.

In this study, I examined how third grade students in the state of focus in this study performed on the standardized mathematics assessment based on their enrollment in STEM and non-STEM schools from 2012 to 2017. The standardized measurement instrument was administered in 2012 in all state K-12 public schools. Since the instrument was a new design administered initially in 2012, the state educational agency phased in the passing requirements by increasing the number of test items students needed to answer correctly over time from 2012 until 2016. This extended phase-in method provided students and teachers necessary time to adjust to the rigor of the exam. The state contracted with Human Resources Research Organization (HumPRO) based on the house bill from the state legislature, HB743, which mandated that the assessment instrument be empirically vetted for validity and reliability by an independent organization before being administered to students. The state education association met the empirical evidence standard by establishing three tasks. Task 1 was to identify that the contents on the mathematics instruments were valid by rating the sufficiency of each test item to the state expectations that it meant to measure. Task 2 was to ensure that the projected reliability and conditional standard error of measurement (CSEM) estimates were acceptable. In addition, Task 3 was a review of the procedures followed to construct the instrument was and the methods established by the state education association to score the instrument, which the state education association found to be consistent with industry standards of validity and reliability in test construction.

The state education association established three criteria to analyze for validity. First, Grade 3 standardized mathematics scores needed to represent each student's knowledge and mathematics fluency, which would signify an alignment between grade level essential knowledge and skills expectations and the instrument. Second, the third grade mathematics scores should indicate the level of student knowledge gain when compared to test scores from the previous year to interpret growth between grade levels. Third, the third grade mathematics scores should indicate student potential achievement levels on future tests. The state education association deemed that validity evidence for the second theme, interpreting growth, was out of the scope of review since third grade is the first year of the mathematics administration meaning that no comparison is available as there are no second grade mathematics scores to determine student growth in knowledge gain during the first interval. The third theme, anticipated growth rates, was also determined to be out of the scope for review because the state education association only provides values from standardized test progress measures starting from Grade 4, which is compared to Grade 3.

HumPRO trained its reviewers to rate every test item in categories of (a) fully aligned, (b) partially aligned, and (c) not aligned. Three to four reviewers assessed the test form, and the final ratings were an average of the results from the reviewers at the state education association. A fully aligned rating indicated that each test item fully connected within the set of essential knowledge and skills expectations upon which the test item was based. Partially aligned meant that some of the content standards did not meet the content standards, and not aligned indicated that the test question fell outside of the contents within the essential knowledge and skills expectations.

According to the state education association, HumPRO found that overall ratings linking the essential knowledge and skills to instrument were highly positive. HumPRO reviewed the 2016 Grade 3 mathematics instrument, and three of the 46 items were rated as partially aligned, with the remaining 43 items rated as fully aligned. HumPRO evaluated each of the four categories in the essential knowledge and skills and found that numerical representations and relationships was 92% aligned with readiness standards, and the last three categories including computations and algebraic relationships, geometry and measurement, and data analysis and personal financial literacy were in full alignment with the readiness standards. After rigorous examination of the testing instrument, HumPRO reported that educational association testing processes and scores were valid and reliable.

Data Analysis Plan

IBM SPSS Statistics version 24.0 is the software I used to prepare the data for analysis. The data, which was archived by the educational agency in the state of focus, was publicly available. I organized the data by school type, STEM and non-STEM, and each school will be assigned a unit number. The data were further organized according to the years tested (2012 to 2017) for each school, whole school average scale score for each school, the number of students tested in each school, gender, and predominant ethnicity of students within each school. I imported these data into a Microsoft Excel spreadsheet. The RQs and hypotheses were analyzed by using a longitudinal, retrospective method using IGC models in order to determine any growth trajectories. Shek and Ma (2011) stated that using IGC models are increasingly used as an analytic tool to capture individual change over time.

STEM schools were cluster-sampled and non-STEM schools were stratified sampled from 13,755 third grade students attending 21 STEM schools and 138 non-STEM schools during the academic years of 2012 through 2017. The independent variables, comprised years tested (time) related to school type (STEM, non-STEM), were analyzed against any rates of change between the dependent variable of mathematics test scores, described using inferential statistics from longitudinal data and IGC techniques. IGC models demonstrated any systematic changes in mathematics test scores within STEM schools and within non-STEM schools. IGC modeling also revealed any differences between-school mathematics scores over time from 2012 to 2017.

I analyzed intraindividual and interindividual differences in growth over time, given the results from third grade student standardized mathematics test. In order to accomplish that, two levels of IGC modeling were used. Level 1 model was used to analyze RQ1 and Level 2 modeling was used to answer RQ2. To plot the IGC in SPSS, Version 24.0, the data was converted from “wide” format to “long” format. Two steps or models of IGC analysis were initially used in this study. The first model involved constructing the unconditional mean model, which is a one-way ANOVA model, using input commands in SPSS in order to assess the amount of outcome variation that exists in both intraindividual and interindividual levels. Interindividual differences over time can be determined by the intercept and the slope; therefore, the second model involved constructing the unconditional linear growth curve model using SPSS commands to determine the slope and intercept parameters, which determined if the linear growth rate was constant over time.

A negative slope indicates decrease, and a positive slope indicates increase, while zero indicates constancy. The intercept value (time) gives the initial status of the dependent variable. The Estimates of Fixed Effect (*p*-value) output from SPSS was run to determine if the slope was significant. If the *p*-value was less than .05 then the slope was significant and the variability of the parameters could be explained by interindividual predictors. If there was no interindividual difference in trajectory over time the slope could not be considered statistically significant. In this case, there would be no need to perform further growth curve modeling analysis. However, to test for a nonlinear individual growth trajectory across time, other higher-order polynomial trends, including quadratic and cubic slope models could have been included (Shek & Ma, 2011). The results of this study were interpreted using a confidence interval of 95% and the *p*-value was considered statistically significant at 0.05.

Given the above explanations, I described the rate of change over time in third grade student achievement in mathematics test scores using the following basic linear growth models:

Level 1 Model: $Y_{ij} = \beta 0j + \beta 1j$ (Time) $+ rij$

Level 2 Model: $Yij = \gamma 0i + \gamma 1i$ (Time) $+ \gamma 2i$ (Time2) $+ \gamma 3i$ (Time3) $+ \gamma 4i\ Wj + rij$

The level-one model was developed as shown below using the inserted variables to test the first RQ. The model enabled me to examine any significant variation within individual school changes over time, and to assess any outcome variations across individuals. The level-two model was developed as shown below using the inserted variables to test the second RQ. The model permitted me to examine any significant variation between individual school changes over time.

Level 1 Model (Measures within Individual School Change over Time)

The formula for this analysis is:

$MATHij = \beta_{0j} + \beta_{1j}$ (Time) $+ r_{ij}$

where MATH*ij* is an individual school average STANDARDIZED TEST score at TIME*i*; β_{0j} is the expected estimation of the MATH score for an individual school at TIME zero; β_{1j} is the average annual rate of change in estimation of the MATH score for an individual school over time; and r_{ij} is the residual within the outcome variable for an individual school at TIME.

Level 2 Model (Measures between Individual School Change over Time)

The formula for this analysis is:

$MATHij = \gamma 0i + \gamma 1i$ (Time) $+ \gamma 2i$ ($Time^2$) $+ \gamma 3i$ ($Time^3$) $+ \gamma 4i\ Wj + rij$

where MATH*ij* is the grand mean for the STANDARDIZED TEST scores for the whole sample at TIME*t*; γ0*i* is the initial average STANDARDIZED TEST score for the whole sample at TIME*t*; γ4*i* tests if TIME is associated with growth parameters; W*j* measures the effect of TIME on interindividual variation on MATH scores; and *rij* refers to the amount of variance that are unexplained by TIME.

Statistical Programming

According to Shek and Ma (2011), the following syntax can be programmed into SPSS to perform an analysis for the unconditional linear growth curve model. I ran the following program developed by Shek and Ma (2011) to test the unconditional mean model for non-STEM schools:

```
mixed Math_Average_Scale_Score_NS with YearTraditional
/fixed intercept YearTraditional
/random intercept YearTraditional | subject(Unit_ID_Traditional) covtype(un)
/print solution testcov /method ml.
```

More detailed programming commands connected to the syntax for the non-STEM school analysis as shown above and their interpretation are illustrated in Appendix A.

I used the following syntax developed by Shek and Ma (2011) to test the unconditional mean model for STEM schools:

```
mixed Math_Average_Scale_Score_S with YEARSTEM
/fixed intercept YEARSTEM
/random intercept YEARSTEM | subject(Unit_ID_STEM) covtype(un)
/print solution testcov /method ml.
```

More details regarding the programming commands connected to the syntax for the STEM school analysis as shown above and their interpretation are illustrated in Appendix B. Through this analysis I expect to be able to determine differences between mathematics achievement of children enrolled in the primary grades in STEM and non-STEM schools, and also change over time within STEM and non-STEM schools regarding children's mathematics achievement.

Threats to Validity

According to Yu and Ohlund (2010), different types of external and internal validity threats exist, and particular factors might cause potential problems in data interpretation; therefore, the design of the research is critical and must be considered in order to minimize potential threats. External validity is important in quantitative research, because it determines whether findings from a research study can be generalized to other populations (Creswell, 2013; Yu & Ohlund, 2010).

The educational agency minimized potential threats to external validity by addressing particular design aspects of the instrument. The assessment instrument is reliable and valid. There were no obvious threats to validity because of the structure in which the educational agency established the instrument. The educational agency minimized threats to validity through its processes as described in detail earlier in this Instrumentation section. No generalizations were established beyond the bounds of the sample population in order to avoid threats to external validity. Because data were preexisting, there were no internal validity threats.

Ethical Procedures

This study was conducted following approval from the Walden Institutional Review Board (approval # 11-07-18-0024471). In order to adhere to ethical procedure, all participants in this study were protected as they remained anonymous and not used for any economic or personal gain. The school district that is the focus of the study was also protected in that its name will remain anonymous. The data used in this study were and are publicly available, and student names are not linked to test scores or any other personally identifiable information. The names of the schools are listed in the analytic portal, but not mentioned herein. I examined the records of the school district's third grade students. The school district that was the focus of this study is different from my own work environment and the grades under study were different from the grade level that I teach.

Summary

The retrospective, longitudinal approach using IGC models for this study allowed me to determine whether mathematics scores from third grade student state-mandated standardized test differed between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. By examining state-mandated standardized mathematics test scores of third grade students using growth curve modeling, I determined if there were any within-school or intraindividual differences in the growth trajectory over time, or if there were any interindividual differences between schools over the six-year time period observed in this study from 2012 to 2017.

I used the Longpower package to calculate the appropriate sample size for longitudinal data, and the methodology was be a retrospective, longitudinal IGC model to analyze and interpret the results. In Chapter 4, I will present the statistical analysis based on the RQs, and I will explain and interpret the results.

Chapter 4: Results

The purpose of this quasi-experimental study using retrospective, longitudinal data and IGC models, which comprises two levels of analysis (Level 1 and Level 2), was to determine whether mathematics scores from third grade student state-mandated standardized mathematics test differ between students who were enrolled in STEM-based schools and students who were enrolled in non-STEM schools. The two categories or domains that were measured in this research study included standardized mathematics test scores of third grade students enrolled in STEM schools and standardized mathematics test scores of third grade students enrolled in non-STEM or traditional schools in an urban public school district in a Southwestern state of the United States. The data represented in my study were publicly available. The sample included third grade student average mathematics scores from the annual state-mandated standardized test, which were examined longitudinally across six different time periods from 2012 to 2017. The RQs and hypotheses that guided this study were:

RQ1: What are the individual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$1: There are no statistically significant changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$1: There are statistically significant changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

RQ2: What are the between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$2: There are no statistically significant differences in between-person or interindividual changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$2: There are statistically significant differences between-person or interindividual changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

This chapter includes an overview of (a) the data collection process that I used to analyze each RQ, (b) baseline descriptive statistics, (c) demographic characteristics, and (d) data analyses procedures I used to address the statistical assumptions of the study to determine whether the underlying requirements of the analyses performed were met. I then present the results from the statistical analyses and provide justifications based on the analyses of the sample to demonstrate whether any interindividual and intraindividual changes over time in average mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools exists.

Data Collection

The data for this study comprised multiyear state-mandated standardized mathematics test scores of third grade students enrolled in STEM schools and third grade students enrolled in non-STEM schools in the largest urban public school district in a Southwestern state. I followed a systematic process to extract the variables that define the data, which were publicly available to me on the Data Interaction Page for Student Assessments, the assessment arm of the educational agency in the state of focus in this study. I collected data at different measurement points including Spring 2012, Spring 2013, Spring 2014, Spring 2015, Spring 2016, and Spring 2017. No missing data were reported. Of the 159 elementary schools in the target district, 21 are STEM schools, and 138 are non-STEM schools. Two elementary schools were ineligible because they are separately a part of the district’s elementary and middle school combination, meaning they are housed in the same educational facility as a middle school. The third school was excluded because it was the single school out of 159 schools not part of the Title I program that supports achievement in high-minority, low income areas (see Kainz, 2019), and so its student population may have been distinct from the populations in the remaining schools. The three schools that were not included in my final analysis finalized the sample size of STEM schools at 18. I limited my sample of non-STEM schools to 18 to ensure a balance in the number of eligible STEM schools.

To achieve a balance in the number of 18 STEM to 138 non-STEM schools, I used the R Project for Statistical Computing (R) software. Longitudinal studies can be designed using balanced and unbalanced data (Shek & Ma, 2011), but an unbalanced design is considered incomplete (Laird, 2004). I created a balanced design because all individuals (*n* = 216) were measured at the same occasions from 2012 through 2017. Based on an assumption of IGC models, balanced data across different observation years of data is not necessary. However, when possible, using an equal sample size is suitable, as it ensures the study has larger statistical power, is less susceptible to homoscedasticity, and is complete, and it facilitates analysis and interpretation (Laird, 2004; Shek & Ma, 2011).

Once I used R to generate a random set of 18 schools from a list of 138 non-STEM schools, I followed the same systematic process to generate the report for the non-STEM schools from the Data Interaction Page for Student Assessments Portal: Assessments, the assessment arm of the educational agency of the state of focus in this study. Reviewing individual average scale scores of each school over time, including examination of the average scale scores, took approximately 2 weeks to complete.

Demographics

The school district of focus in this study describes its enrollment policy as a district of choice. It does not recognize attendance zones for its specialized schools, which are all a part of its magnet programs. Entry into the district STEM schools is based on a three-phase application timeline it devised for students who meet eligibility guidelines for the program. The application process is opened to any registered student who lives within the district's boundaries and also to children of active school district employees. If the program has more applicants than there are spots, student names are entered into a lottery system. Once eligible students are placed in the STEM schools, students who are considered out-of-district are placed in STEM schools as space is available. Students who attend non-STEM schools in the district and do not desire to apply for transfers to specialized schools must enroll in the school zoned to their homes. Excluding the 13 students over time who did not identify a gender, the number of participants who took the state-mandated mathematics standardized test is recorded in Table 3. While gender or ethnicities were not foci of the study, the data was publicly available and recorded when data was collected and is included as a part of the demographics of the sample. The majority of the students in this study who took the mathematics standardized test were female (52%, *n* = 36) to male (48%, *n* = 36) also shown in Table 3.

Table 3.

Number of Participants at Each Measurement Occasion

Year	Spring 2012	Spring 2013	Spring 2014	Spring 2015	Spring 2016	Spring 2017
N(School)	36	36	36	36	36	36
Number of participants	11090	11053	12136	12657	13322	13755
Male	5575	5560	6092	6400	6670	7017
Female	5509	5493	9044	6257	6651	6735
No gender reported	6	0	0	0	1	6

Results

Descriptive Statistics

The Level 1 stage of analysis of the individual growth trajectory over time of overall average mathematics test scores of the non-STEM school participants (*M* = 1431.81, *SD* = 53.734, *n* = 108) compared to the overall average mathematics test scores over time of STEM school participants (*M* = 1431.17, *SD* = 51.665, *n* = 108) revealed that there was no distinguishable difference between individual test scores. The relationship between the STEM-based mathematics instruction and the shape of each STEM school individual growth trajectory over time compared to the relationship between each non-STEM school mathematics instruction and the shape of each non-STEM school individual growth trajectory over time indicates that there is no difference between mathematics test scores; therefore, the findings were nonsignificant. Higher standard deviations indicate greater levels of performance inconsistency in relation to mean scores. Based on the higher standard deviations in both the STEM and non-STEM scores, there were greater levels of performance inconsistencies as shown in Figure 1.

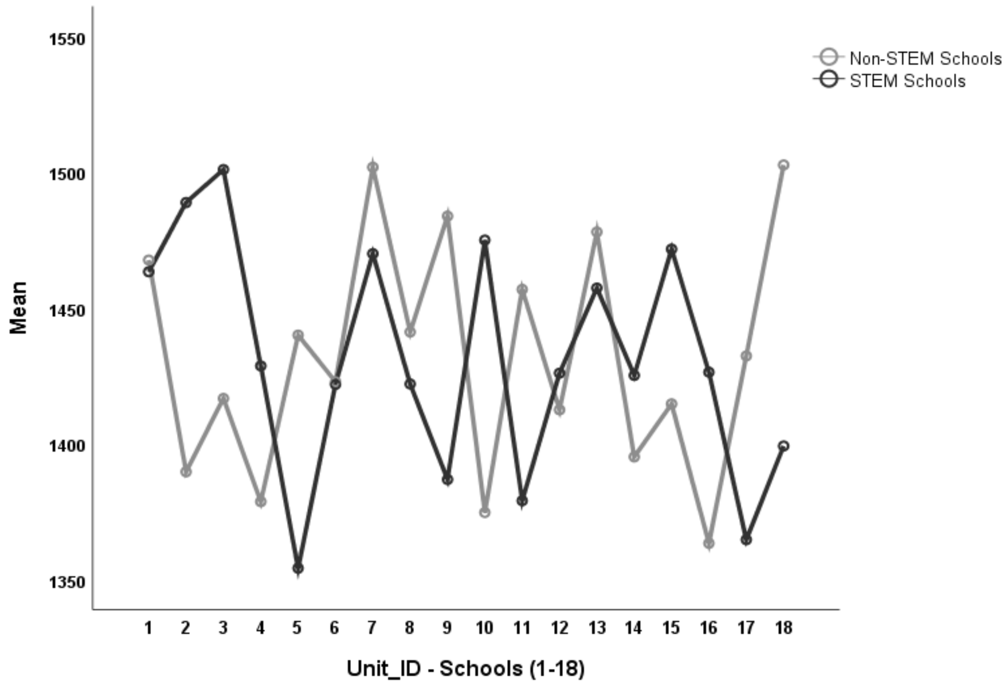


Figure 1. Mean mathematics scores of non-STEM and STEM schools over time.

Individual group mean scores over time, the overall average of each domain, and standard deviations from non-STEM schools and STEM schools are shown in Table 4.

Table 4.

Mean Test Scores for Non-STEM and STEM Schools over Time

Mean mathematics test scores between groups	Year	*N*	Mean	Standard deviation
Non-STEM mathematics mean score	2012	18	1430.89	54.051
Non-STEM mathematics mean score	2013	18	1421.28	59.522
Non-STEM mathematics mean score	2014	18	1434.28	46.729
Non-STEM mathematics mean score	2015	18	1425.22	48.704
Non-STEM mathematics mean score	2016	18	1433.22	56.432
Non-STEM mathematics mean score	2017	18	1445.94	59.866
Overall mean test score – non-STEM group		108	1431.81	53.734
STEM mathematics mean score	2012	18	1428.56	34.104
STEM mathematics mean score	2013	18	1431.94	55.847
STEM mathematics mean score	2014	18	1442.11	57.562
STEM mathematics mean score	2015	18	1421.83	56.096
STEM mathematics mean score	2016	18	1433.28	54.621
STEM mathematics mean score	2017	18	1429.28	53.290
Overall mean test score – STEM group		108	1431.17	51.665

Data Analysis Procedure

Model Building–Level 1 and Level 2. I analyzed the data by using a mixed-effect model with maximum likelihood estimation (MLE), as MLE is flexible and most appropriate when handling real data. This method modeled individual changes over time, determined the shape of the growth curves, explored systematic differences in change, and examined the effects of predictors in the initial status and the rate of growth. This is an appropriate approach in the study of individual change because it creates a two-level hierarchical model that nests time (year) within individuals. There are two levels in IGC models. The Level 1 model in this study encompasses Equation 1 and Equation 2, answers the first RQ by describing within-individual or intraindividual changes (i.e., repeated measures) over time. Level 1 focuses on the individual school average mathematics test scores and describes each one's developmental changes or variations over time. The Level 2 model, which is modeled in Equation 3, answers RQ2, and captures whether the rate of change varies across individuals in a systematic way. Basically, it describes any variation related to the interaction between the population samples. The growth parameters such as the within-subjects intercepts and slope of Level 1 (RQ1) are the outcome variables to be predicted by the between-subjects variables at Level 2. Because of the complexity of the IGC model, two outside statisticians were asked to and did confirm the analysis and my presentation of the results.

Level 1 Model–Equation 1

The Level 1 model of analysis answers the first RQ, which was:

RQ1: What are the individual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$1: There are no statistically significant changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$1: There are statistically significant changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

The Level 1 model represents intraindividual or within-school changes that each school is expected to experience from the initial status (Year 1 = 2012) and the rate of change over time (2013 through 2017). No predictors are included in the Level 1 model, as it focuses strictly on outcome values, which are time-variant. The outcome values in the Level 1 model are time-variant meaning that the growth trajectory depend explicitly on how the scores change with time. There are two equations in the Level 1 model, Equation 1 (1) and Equation 2 (2). The basic linear growth model, Equation 1, is described below:

Level 1 Model (Equation 1):

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Time}) + r_{ij} \qquad (1)$$

In this model, Y*ij* is the repeated measurement of average mathematics test scores for an individual school *i* at Time *t*, where β0 is the initial status, the first year of the longitudinal trajectory (Year 1 = 2012) of the average mathematics test scores for individual schools *i*, and where *j* represents each observation year (2012 through 2017). β1j is the linear rate of change for individual schools *j*, and *rij* is the residual in the outcome variable y for individual schools *j* at Time *t*. The residual is the difference between the observed *y*-value and the predicted *y*-value for a given *x*-value on the regression line. For example, if the predicted score from my model were 1500, then *rij* = (observed *y*-value of 1471) – (predicted *y*-value of 1500). RESIDUAL *i* =1, 1 being year 2013, *j* = 3, 3 being the name of the school, (School 4) then (observed *y*-value 1471) –

(predicted *y*-value 1500) = 29. The residual variance determines whether there is linear rate of change or nonlinear rate of change. If the effect of linear growth (Time, β1) is not statistically significant, there is no need to perform further growth curve modeling analysis.

Level 1 Model–Equation 2

To test a nonlinear individual growth trajectory across time, other higher-order polynomial trends including quadratic and cubic slopes can also be used for model testing, which is shown in Equation 2 below:

Level 1 Model (Equation 2) was:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Time}) + \beta_{2j}\,(\text{Time}^2) + \beta_{3j}\,(\text{Time}^3) + r_{ij} \qquad (2)$$

In Equation 2 of the Level 1 Model, *Time* in the linear slope, β_1, remains in the equation, while $Time^2$ in the quadratic slope, β2, and $Time^3$ in the cubic slope, β_3, are added. The linear slope suggests that the rate of growth remains constant across time and is represented by a straight line. Higher-order polynomial trends indicate that the rate of growth may differ over time. The quadratic individual change trajectory, the second-order polynomial, has a curved line and no constant common slope as the data can fluctuate between gains and losses over time, and consists of a single stationary point including a peak and trough. A cubic trajectory has two stationary points with one peak and one trough that is S-shaped.

Level 2 Model–Equation 3

The Level 2 model of analysis answers the second RQ:

RQ2: What are the between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools?

$H_0$2: There are no statistically significant differences in between-person or interindividual changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

$H_1$2 – There are statistically significant differences between-person or interindividual changes over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools.

On the Level 2 model, which is represented by Equation 3, an explanatory variable (W*j*) would be included to analyze the predictor's effect on interindividual variation on the outcome variable. When a variable is not completely independent, it is explanatory in that it offers additional explanation for patterns of change in individual growth trajectory (Singer & Willett, 2003). The errors are assumed to be independent and normally distributed, and the variance is equal across individuals. The Level 2 model, Equation 3 (3), is shown below:

Level 2 Model (Equation 3) was:

$$Y_{ij} = \gamma_{0i} + \gamma_{1i}\,(\text{Time}) + \gamma_{2i}\,(\text{Time2}) + \gamma_{3i}\,(\text{Time3}) + \gamma_{4i}\,W_j + r_{ij} \quad (3)$$

In this equation, Y_{ij} is the grand mean for the mathematics test scores for the whole sample at Time *t*. γ_{0i} is the initial status of the mathematics test scores for the whole sample at Time *t*. γ_{1i} is the linear slope of change relating to the mathematics test scores for the whole sample at Time *t*. γ_{2i} is the quadratic slope of change relating to the mathematics test scores for the whole sample at Time *t*. γ_{3i} is the cubic slope of change relating to the mathematics test scores for the whole sample at Time *t*. γ_{4i} is used to test whether the predictor (e.g., group) is associated with the growth parameters (i.e., initial status, linear growth, quadratic growth, and cubic growth). Random effects (i.e., amount of variance) that are unexplained by the predictor are referred to as r_{ij}.

Step 1: Unconditional Mean Model (Model 1)

IGC modeling was used to examine the individual growth trajectory of each school can be examined in the empirical growth plot of each school, which is found in Appendix C and Appendix D. Since this step focused only on the patterns of change in test scores over time, there is no predictor included in it. This step serves as a baseline model in the outcome variable without regard to time. This model assesses (1) the mean of the outcome variable and (2) the amount of outcome variation that exists in intraindividual and interindividual levels. This latter information is important as it helps determine which level (i.e., Level 1, time-variant or Level 2, time-invariant) of predictors to add when fitting the subsequent models. If the variation is high, it suggests that the predictors at that level could explain certain amount of outcome variation. According to Shek and Ma (2011), one of the strengths of the IGC model is that it examines the proportion of total outcome variation that is related to interindividual differences (i.e., intraclass correlation coefficient [ICC]). The ICC describes the amount of variance in the outcome that is attributed to differences between the STEM schools and the non-STEM schools. It evaluates the necessity of modeling the nested data structure (i.e., any significant variation in individual initial status of the outcome variable). It is also a measure of the average autocorrelation of the outcome variable over time, meaning it is the expected correlation between any two randomly chosen schools in the same group (Heck et al., 2014).

The higher ICC value indicates the estimated average stability or consistency of the dependent variable over time within groups, meaning that a substantial variance indicates that the groups are relatively homogeneous, which determines that they are likely highly different from each other (Heck et al., 2014). Stability or instability of test scores over time has important implications for establishing effective policy regarding potential factors that influence patterns of change. ICC values range from 0 to 1. When an ICC value is close to 1 it is considered a higher value, which indicates a high similarity between test scores from the same group. When the ICC value is low, which will be close to zero, it reveals that the values within the same group are not similar.

Research Question 1 results: Individual intraindividual changes in growth over time within the non-STEM schools average mathematics test scores were non-significant ($p = 0.09$). Individual intraindividual changes in growth over time of the STEM average mathematics test scores were non-significant ($p = 0.07$). Based on the results there are no statistically significant changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. Results showed that when the STEM and non-STEM average mathematics standardized test scores of third grade students in the school district of focus were compared, the growth trajectories were statistically nonsignificant as seen in Figure 2, and illustrates that there was no statistical difference in test scores over time between the STEM group and the non-STEM group.

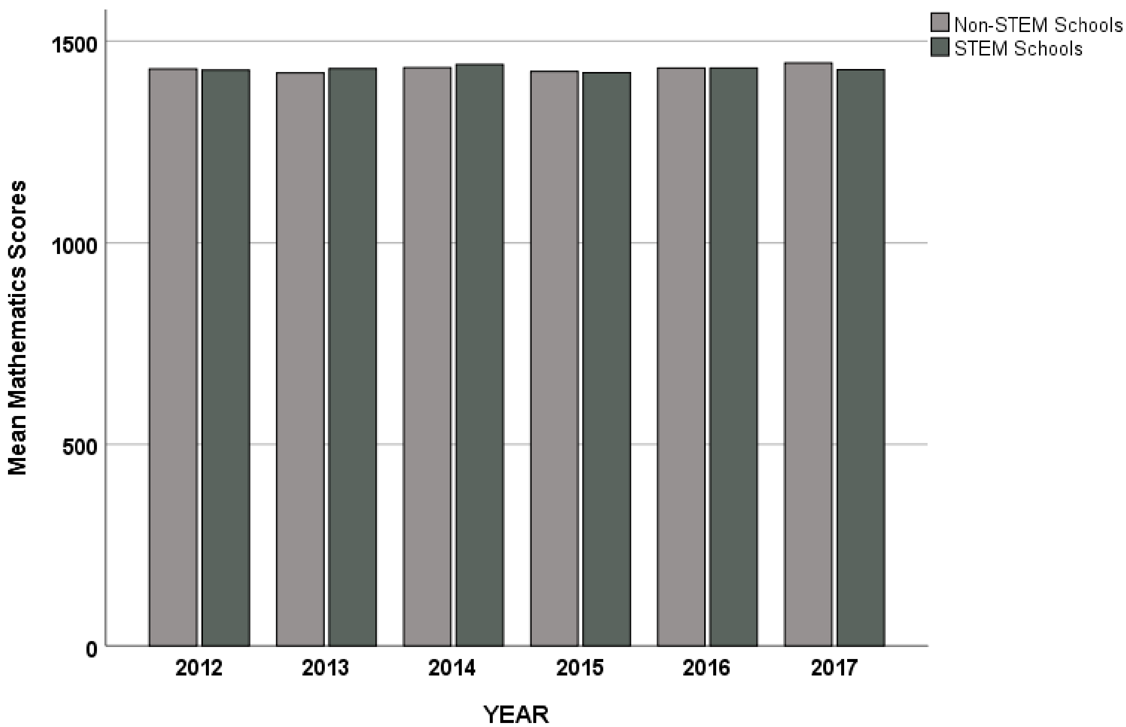


Figure 2. Mean outcome values of STEM and non-STEM mathematics test scores.

Koo and Li (2016) and Spybrook, Raudenbush, Liu, Congdon, and Martinez (2006) stated that ICC values that exceed 0.40 are common in longitudinal social research studies. ICC results from the estimates of covariance parameters for non-STEM schools, which can be found in Appendix E, was 1541.54/(1541.54 + 1319.03) = 0.539. This value suggested that about 53.9% of the total variation in the average test scores was due to interindividual differences (see Figure 3). Generally, IGC modeling is required if ICC is 0.25 or above. Given that the ICC for this study is above 0.25 (53.9%) an ANOVA would have been an inappropriate statistical method to use to analyze the data; it cannot answer my RQs. According to Shek and Ma (2011), if the ICC is low, IGC might not perform better than the traditional method (e.g., ANOVA) in estimating fixed effects. The estimated average stability of the average test scores at 0.539 is an alert that there are possible mediating and, or moderating effects on outcome variables.

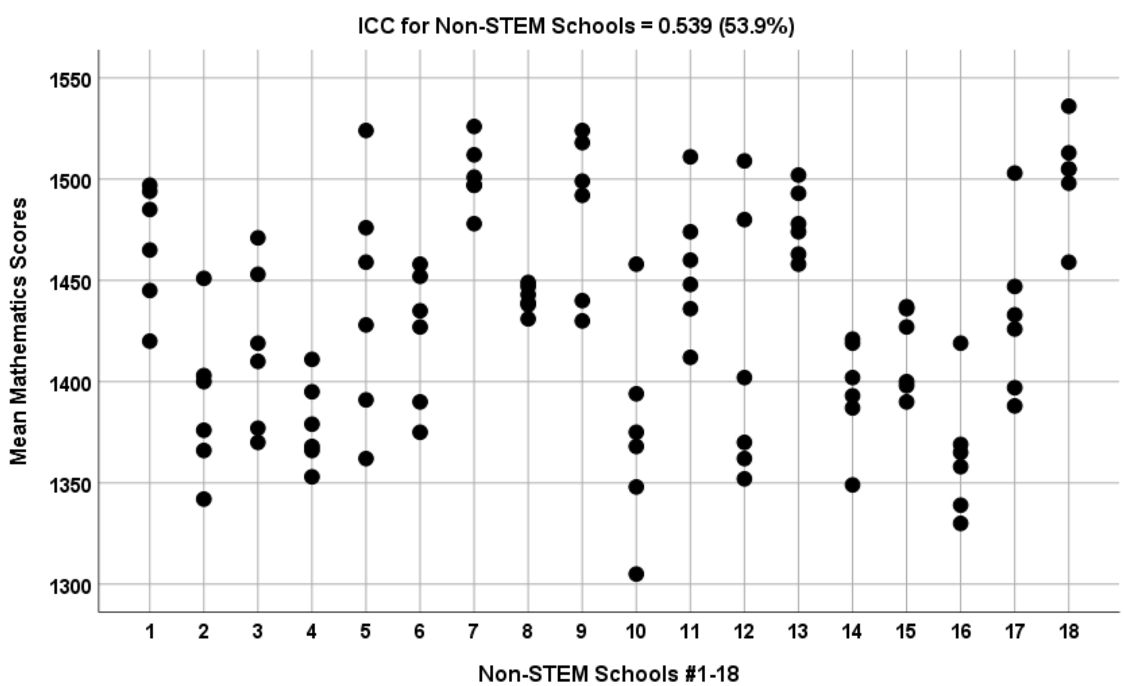


Figure 3. Dotplot of non-STEM population scores.

The ICC for STEM schools was 1559/(1559+1086) = 0.589, as shown in Figure 4, suggesting that about 58.9% of the total variation in the average test scores was due to interindividual differences (RQ2 – between-person changes). The estimated average stability of the average test scores was 0.589. If ICC is low, closer to zero, the IGC might not perform better than the traditional method (e.g., ANOVA) in estimating fixed effects. Given that the ICC for STEM schools is 53.9% and 58.9% for non-STEM schools, which are both higher than 0.25, IGC modeling is required. The higher ICC percentages demonstrate the estimated average stability of the dependent variable over time showing that the non-STEM schools had higher stability in outcome values. Furthermore, it is an alert that there are possible mediating and, or moderating effects on outcome variables.

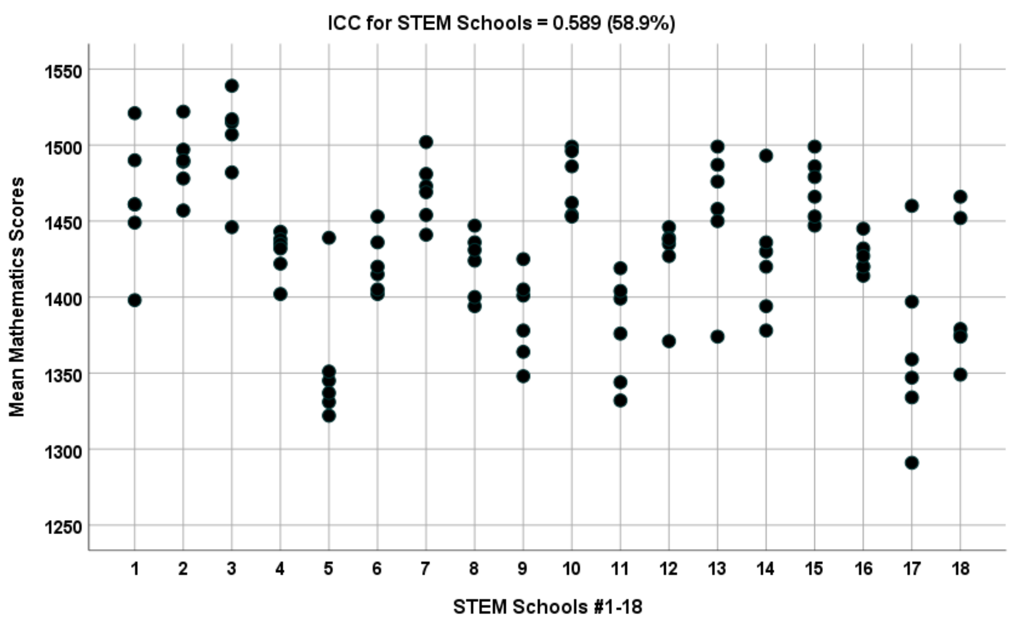


Figure 4. Dotplot of STEM population test scores.

Step 2: Unconditional Linear Growth Curve Model (Model 2)

This model serves as the baseline growth curve model to examine individual variation of the growth rates (i.e., any significant variations in individual trajectory changes over time) and will answer the second RQ. Unlike the unconditional mean model, which only assesses the outcome variation across individuals (i.e., the differences between the observed mean value of each school and the true mean from the population) this model also examines individual changes over time (i.e., how each school rate of change deviates from the true rate of change of the population). If there is no interindividual difference in trajectory change over time (i.e., *Time* is not statistically significant), further model testing would not be performed.

Research Question 2 results: In this study, the non-STEM data shows no interindividual or between-school group (non-STEM and STEM) differences over time, because time was not statistically significant ($p = .308$); therefore, higher model testing such as the quadratic and cubic growth curve models, after Step 2, were not needed. The significant values in both the intercept and linear slope parameters indicate that the initial status and linear growth rate were not constant over time. The mean estimated initial status and linear growth rate for the non-STEM group was -4442.22 (Appendix F). This mean estimate was not significant because the *p*-value was .438. The linear growth rate for the sample was 2.92, and since the linear growth rate trended towards being positive, the non-STEM schools mean test score trended upwards with time. The random error terms associated with the intercept and linear effect were not significant ($p > 0.05$), suggesting that the change in these parameters could not be explained by between-individual predictors, or cannot be explained by interindividual non-STEM differences. Further research that examines other mediating or moderating variables of concern to the target district will be necessary to determine intraindividual and interindividual differences in test scores.

The within-individual changes demonstrate that there were differences within the same schools in the non-STEM group over time. The correlation (β = -164221, *SE* = 95203, *p* =.085, *p* > 0.05) found in Appendix F between the intercept and the linear growth parameter trended towards being negative. This suggests that non-STEM schools with high average test scores trended towards a linear decrease, whereas non-STEM schools with low average test scores trended towards a faster decrease in linear growth over time (see Figure 5).

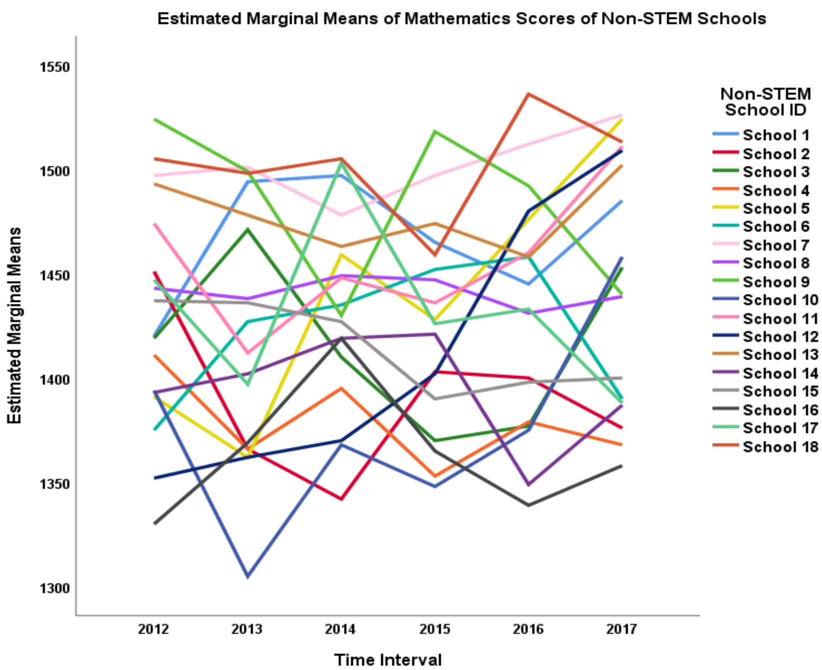


Figure 5. Estimated marginal means of mathematics scores of Non-STEM schools.

The significant values in both the intercept and linear slope parameters of STEM schools indicate that the initial status and linear growth rate were not constant over time. The linear growth rate in the average test scores found in STEM schools trended upwards (β = -.362, *SE* = 2.46, *p* = .885). The mean estimated initial status and linear growth rate for the STEM schools sample was 2160.22. Even though this was the mean estimate for STEM schools, it was not significant because the *p*-value was .885 ($p = .885$, $p > 0.05$). The linear growth rate for the sample was -.362; the values can be found in Appendix G. Since the linear growth rate trended towards being negative, the dependent variable decreased with time. This suggested that the mean score for STEM schools was 1431.17, and the growth trajectory showed that it trended upwards with time. The random error terms associated with the intercept and linear effect were not significant (*p* >. 0.05), suggesting that the change in these parameters could not be explained by between-individual predictors, or cannot be explained by interindividual STEM differences.

The suggestion is that there are intraindividual differences in STEM schools. There are differences within the same schools in the STEM group over time (see Figure 6). Unexplained differences in individual growth parameters suggest that multiple related factors exist that can explain the variability. The correlation (β = -118956, *SE* = 75252, $p = .114$, $p > 0.05$) between the intercept and the linear growth parameter trended towards being negative (see Appendix G). This suggests that STEM schools with high average test scores trended towards a slower linear decrease, whereas STEM schools with low average test scores trended towards a faster decrease in linear growth over time.

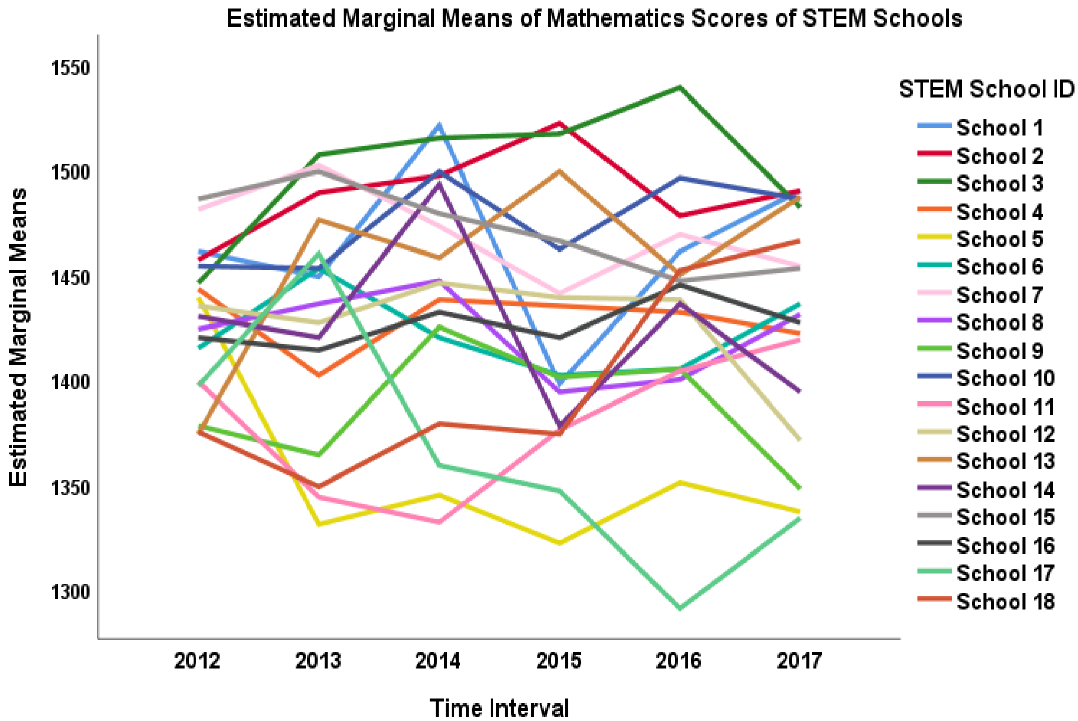


Figure 6. Estimated marginal means of mathematics scores of STEM schools.

Additional findings: The largest ethnic group within the participants in each year in the sample was Hispanic or Latino with a steady increase of participants from 5524 in the initial year (2012) to 7420 in the last time period (2017). Further information on each ethnicity in the study sample is in Table 5. The American Indian, Native Hawaiian, and two or more participants were in single and double digits. Fewer than 3% of individuals from each year did not provide an ethnic background. Asian students had the highest mathematics test scores in each observation year. Of the students who provided an ethnicity in each year, the African American students had the lowest mathematics test scores. The study sample is representative of the population of interest, and proportional, given that I used probability sampling to determine the non-STEM population.

Table 5.

Number of Participants by Ethnicity at Each Measurement Occasion

	Spring 2012	Spring 2013	Spring 2014	Spring 2015	Spring 2016	Spring 2017
Hispanic or Latino	5524	5534	6122	6661	7052	7420
Black or African American	3655	3530	4094	4086	4160	3996
White or Caucasian	1268	1225	1207	1172	1259	1348
Asian	491	489	496	529	592	749
Two or more races	100	141	171	165	178	197
American Indian or Alaskan Native	26	28	20	20	20	20
No ethnicity provided	17	86	7	6	50	15
Native Hawaiian or Other Pacific Islander	9	20	19	18	11	10

Summary

This research document describes individual growth trajectories of standardized mathematics test scores over time of third grade students who were enrolled in 18 STEM schools compared to third grade students who were enrolled in 18 non-STEM schools in the same school district. The study was guided by two RQs that sought to predict within-individual changes in growth over time and between-individual variability in growth of outcome values from third grade student standardized mathematics test scores based on their learning experiences in STEM-based mathematics instruction and non-STEM mathematics instruction. In addition to time (year), which is considered as an independent variable in growth analysis, STEM-based mathematics instruction and non-STEM mathematics instruction were predictors of change used to analyze systematic variation in growth trajectories over time. Based on the results from the individual growth patterns the average mathematics scores over time between the STEM and non-STEM schools were statistically insignificant as related to RQ1.

The linear growth rate of the non-STEM schools was not constant over time, and there were differences within the same schools in the non-STEM group over time. The linear growth rate of the STEM schools as related to RQ2 trended towards being negative. The fluctuations in the growth trajectory over time were not significant, which implies that the growth patterns in the scores cannot be explained by the between-school predictors, but possibly by further researched of multiple covariates. Demographic results displayed that the majority of the third grade students in this study who took the mathematics standardized test were female. The demographic composition of the sample showed that Latino or Hispanic students represented a higher percentage than African American, White, and Asian students. In Chapter 5, I will present an interpretation of these findings, along with implications for social change, recommendations for action, and recommendations for further study.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this quantitative study was to determine whether mathematics scores from third grade student state-mandated standardized mathematics test scores differ between students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. In this quantitative, retrospective, longitudinal study of change, I used a simple two-level growth model that encompasses Level 1 and Level 2. The Level 1 model focused on within-individual patterns of change over time, and it is these patterns that characterized each school's individual growth trajectory over time. The Level 2 model asks what predicted the variability in the growth rates from the STEM schools and the growth rates from the non-STEM schools that were produced from the Level 1 model, as well as any explanations for the patterns of within-individual change over time between each group (STEM and non-STEM) and within each individual school. I used IBM SPSS version 24 to analyze the data and generate results. The sample in this study included 18 STEM schools balanced with 18 non-STEM schools. The sample was gathered from the largest school district located in a Southwestern state, which is the only district in the vicinity with a large number of dedicated elementary STEM schools in addition to its body of 138 non-STEM elementary schools.

Key findings from the Level 1 model demonstrated that the IGC from each school was nonlinear. However, the group growth curve, which included the weighted mean from STEM schools (M = 1431.17) and the weighted mean from the non-STEM schools (M = 1431.81) was not significantly different over time. In examining empirical growth records of within-school individual outcome values over time, almost every school's results fluctuated, some more significantly than others. Key findings from the Level 2 model showed that there were between-school variations in growth rates over time of mathematics test scores but left unexplored possible variables of gender, ethnicity, teacher efficacy, and ecological or contextual factors and how these may have influenced discontinuity observed in test scores over time as demonstrated in Level 1. In this chapter, I further discuss the implications of the findings, the limitations of the study, recommendations for further research, recommendations for practice, implications for positive social change, and conclusions.

Interpretation of Findings

Research Question 1

RQ1 asked about the individual changes in growth over time of mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. Key findings from RQ1 demonstrated that: (a) the average mathematics scores over time between the groups, STEM and non-STEM schools, was statistically insignificant; (b) within-school growth trajectory over time of the STEM schools and within-school school growth trajectory over time of the non-STEM schools was not significant; and (c) the ICC for STEM schools was five percentage points lower than that of the ICC for non-STEM schools, which meant that the estimated average scores over time of the non-STEM schools had higher stability than the estimated average scores over time of the STEM schools.

These findings were not consistent with current literature that indicated that a STEM-based approach to mathematics instruction in the target state of focus in this study could have a positive effect on standardized test scores and increase the possibility of mathematical achievement (McDonald, 2016). Singer and Willett (2003) stated that significant variance in individual growth parameters on Level 1 analysis indicates the influence of possible covariates. My RQ focused singly on individual growth patterns over time to determine the effects of STEM and non-STEM education on standardized mathematics test scores using IGCs. Additional covariates were not included as supplementary questions in the RQ structure but would serve the district well as future research in its quest to provide top rate education for all. It is the nature of growth curve analysis to first determine individual growth trajectories before including explanatory variables to clarify the intraindividual and interindividual differences.

While traditional methods of teaching have some advantages for student learning, mathematics instruction that is STEM-based aligns with tenets found in Polya's (1957) heuristics of problem-solving, which are promoted by NCTM (2000) and other educational stakeholders. As described in the literature review, both STEM and non-STEM disciplines rely on talented workers with STEM-related skills to accomplish job-related tasks in computation, programming language, and digitalization as a means to keep pace with technological trends and innovation. Because STEM-based education typically is student-centered, project-based, and hands-on, some research describes it as more relevant to students than are traditional methods (NCTM, 2000), more connected to real-world contexts, and more motivating for young learners. Government agencies, educational associations, school districts, and other stakeholders believe in the relevancy of STEM education as a more modernized and sensible approach to student learning, which is most effectively achieved with teacher facilitation. As a result, many of the above-named agencies and stakeholders are now implementing STEM-based teaching and learning in some manner in their professional paradigms. The finding of RQ1, that no difference in mathematics achievement occurred for students taught in STEM schools compared to those taught in non-STEM schools, suggests that STEM education in the target district can be embraced wholeheartedly, with confidence in the continued mathematics achievement and the added benefit of infusing fun in a subject that some students find difficult to learn.

Research Question 2

RQ2 asked about the between-person or interindividual changes in growth over time in mathematics scores from a state-mandated standardized test of third grade students who were enrolled in STEM-based schools and students who were not enrolled in STEM-based schools. The Level 2 model analysis, which detects the heterogeneity in patterns of change across schools that were presented from the Level 1 model analysis, links the changes in patterns with the cause, which would be the result of a predictor (e.g. teacher experience, ethnicity, or socioeconomic status). What factors caused fluctuations in growth trajectories were not explored in this study, which is common as an initial step in basic IGC analysis.

Key findings from RQ2 related to non-STEM schools demonstrated that: (a) the differences between test scores that were found at each time period of the non-STEM group were not statistically significant, and therefore, there are no between-school or interindividual differences over time; (b) the outcome values produced in both the intercept and linear slope parameters indicated that initial status and linear growth rate of the non-STEM schools was not constant over time; (c) even though the mean estimated initial status and linear growth rate for the non-STEM group was not statistically significant over time, the group experienced a linear growth rate that trended towards being positive, demonstrating that its test scores increased over time; (d) the random error terms associated with the intercept and linear effect were not significant, which suggests that the change in the parameters could not be explained by between-school predictors, meaning there were no significant effects on the test scores based on the non-STEM curriculum; (e) the change in the parameters over time of the non-STEM scores also cannot be explained by between-school or interindividual differences found between individual growth trajectory of each non-STEM school; (f) the correlation between the intercept and the linear growth parameter in non-STEM schools trended towards being negative, suggesting that non-STEM schools with high average test scores had a slower linear decrease, whereas non-STEM schools with low average test scores had a faster decrease in linear growth over time; and (g) there are differences within the same schools in the non-STEM group over time, which can possibly be explained by researching further with mediating and/or moderating variables.

Key findings from RQ2 related to STEM schools demonstrated that: (a) there are within-school or intraindividual differences over time within the schools in the STEM group, and therefore, potential effects of multiple explanatory, mediating, or moderating predictors could be researched to determine the nature of the differences; (b) the significant values in the intercept and linear slope parameters showed that the initial status and linear growth rate were not constant over time, but the average score over time from the initial status and linear growth rate of the STEM population increased; (c) the linear growth rate trended towards being negative, and therefore, test scores decreased with time, which means that the mean of the STEM group increased over time; and (d) the fluctuations in the growth trajectory over time was not statistically significant, meaning that the patterns of change in the STEM scores cannot be explained by the between-school predictors or by interindividual STEM differences, the STEM-based mathematics instruction approach to learning.

As discussed in Chapter 2, children have a natural affinity for mathematics concepts early on; however, few quantitative studies on the effect of early STEM-based mathematics instruction on third grade student performance on standardized mathematics tests are found in the peer-reviewed literature. Instead, most studies focus on a variety of areas of STEM education predicated on its effect on standardized mathematics tests and mathematics achievement in secondary and postsecondary education (Arik & Geho, 2017; Chiu et al., 2015; Clements & Sarama, 2016; Ejiwale, 2013; McClure et al., 2017; Nguyen et al., 2016; OECD, 2016a). Results of this study that there are no between-school or interindividual differences between mean scores of STEM and non-STEM schools deviates from the general conclusion as reported in the literature review that early STEM education is effective and may positively influence student performance on mathematics test scores.

Summary

My intent in this study was to determine if STEM education and its approaches to learning have an effect on student outcomes on the third grade state-mandated mathematics test. Results of this study indicated that STEM education compared to non-STEM education demonstrated no significant difference on student outcomes on the third grade state-mandated mathematics test. The averages of the between-group (STEM and non-STEM) schools were statistically the same. The results from the empirical growth plots revealed fluctuations in each school individual growth trajectory in which the parameters show similarity in variance between the slope and integers. Such fluctuations indicate that particular predictors other than STEM-based mathematics instruction influenced the interindividual differences in changes over time. Further testing is required to determine what predictors influenced what differences.

Limitations of the Study

The initial intent of this study was to determine if there were statistically significant differences in third grade student standardized mathematics test scores overtime dependent upon STEM and non-STEM pedagogy. In Chapter 1 I reviewed several limitations that may have affected generalizability of findings in this study. Included are limitations due to students who enrolled in STEM programs in the target district with varying levels of STEM exposure in kindergarten through second grade, which may affect mathematics achievement in the third grade in unknown ways. Second, there were no measures available to determine the quality or degree of STEM instruction that students in the sample experienced given the number of classrooms involved and levels of teacher experience. Third, students who live in residential zones different from which the STEM school in which they won lotteries to attend is located are required to transport themselves to school, and this added adjustment for the family may influence student learning in unexpected ways for an unknown period of time.

The following additional findings are not generalizable to all school districts for several reasons: (1) not every school district has established dedicated STEM schools under a specialized magnet program, in addition to traditional, non-STEM schools, as is the case in the school district of focus in this study, (2) there are broad differences in instructional strategies, operational practices, policies, and programmatic implementation in STEM schools across school districts in the state of focus, (3) the research only examined third grade students and not later grade levels to test for longitudinal differences in mathematics achievement, (4) teacher self-efficacy and teacher experience and training levels of education and proficiency in implementing STEM programs may vary widely and might affect student achievement on standardized mathematics test scores, and (5) the target district has experienced administrative difficulties and failure to meet accountability standards, which might affect the application of these findings even in the target district in analysis of more recent data than were included in this study. Therefore, applying these results outside of the scope of this study may be unsuitable.

Recommendations

First, this study provided important empirical data on interindividual variability, which supports further study individual-related factors that account for the variability. Also, the results of the study support a lack of a statistically significant relationship between-school groups or within individual schools, making it reasonable to recommend that a general study on how STEM-based mathematics instruction is designed and delivered in the target district, and in other, similar large school districts. Results of this study dissented from current literature that found a positive interaction between STEM-based mathematics instruction and mathematics achievement, which illustrates the need for expanded, specific, and ongoing research in regards to both the outcomes of this study, and the target district's internal practices and policies regarding its STEM program. In order to best understand the effect of STEM education on mathematics achievement in elementary grades, I recommend that each school district individually analyze outcome measures for its unique population.

The results of this study demonstrated that there is no difference in test scores between STEM and non-STEM schools, indicating that STEM-based education is not academically superior or inferior to traditional education. Given this discovery in the target district, I would recommend that STEM-based mathematics instruction be scaled throughout the district in a STEM-for-all model, which is growing more popular in many school districts and supported by businesses and the federal government. One school district north of the target district has established a partnership with a major company to create a STEM-for-all model as a means to reshape how STEM subjects are delivered and in an effort to make learning relevant for all students. Since the target district in this study has a high population of Title I and economically disadvantaged students who are otherwise under-represented in STEM disciplines and have low enrollment is specialized STEM programs, the STEM-for-all program may provide these at-risk students with an approach to learning that will develop high need skills that are needed throughout their lives and in the workforce (Noonan, 2017; Rothwell, 2013, Sithole et al., 2017).

Trends in individual growth trajectories allow the target district to see patterns of achievement from individual STEM schools, individual non-STEM schools, and between each group. Given the information from the growth patterns, the results suggest that more in depth studies are recommended to pinpoint exact explanations for the intraindividual or within-school variations in test scores, as well as any relationships or interactions between STEM-based mathematics instruction and mathematics achievement, to explain the interindividual differences. For example, because scores varied within schools for unknown reasons, research might focus on factors that possibly caused test scores of STEM and non-STEM schools to increase or decrease over time. Further studies will be necessary to determine what extrinsic factors played a role in the growth trajectories. Factors such as teacher efficacy, teacher accountability, or different teaching styles may have had an effect on student performance that should be considered.

Similarly, future studies might investigate whether differences in growth trajectories found in this study happened because some students learned STEM-based material in earlier grades, or at a faster rate than others, or whether students enrolled in non-STEM approaches to learning retain information at higher rates than STEM students due to the rote memorization practices found in the traditional model. There could be differences due to a student's early mathematics education experiences or STEM exposure from pre-kindergarten through second-grade. Research found that children are mathematically inclined starting at a very young age, which would support this predictor.

Further research might focus more specifically on teaching methods and teacher professional development programs, and their effects on student achievement on mathematics assessments. An experimental study, in which teachers specifically trained in either STEM or non-STEM methods of mathematics instruction, with a comparison of student achievement results, might help elucidate the issue of curricular faithfulness assumed in my study. In addition, in my study I ignored possible teacher differences in self-efficacy regarding mathematics instruction, but this might have been a key factor, because my study compared achievement resulting from primary grade teaching, when school subjects are typically taught by generalists, not by subject matter specialists. Future research, therefore, might explore the effect of teacher self-efficacy in mathematics on student achievement and whether feelings of efficacy vary by STEM or non-STEM curricular model.

This study also identified opportunities for further longitudinal research on the effects of early STEM-based mathematics instructional strategies on children's learning at the end of the primary grades and throughout their middle school, high school, and college careers, and subsequent employment choices. The influence of a district-wide STEM program, from the earliest years through high school graduation, on student learning and careers, is as yet unknown. Since authentic STEM programs provide practical application to real world contexts for learning through hands-on lessons that provide intrinsically appealing opportunities for problem-solving and investigation (Tanenbaum, 2016; Polya, 1957), it is possible that STEM education would result not only in similar achievement to non-STEM, as found in this study, but also increases in student attendance, motivation for school, and graduation rates. Such longitudinal effects should be explored in future research.

Implications

The results from this study showed that there was no significant difference between the average test score between third grade students in STEM schools compared to third grade students in non-STEM schools, meaning that STEM instruction is as effective as non-STEM instruction in mathematics. Therefore, STEM-based instruction can be embraced vigorously and STEM elements may be introduced into more traditional instruction without loss of student learning. Because gains in problem-solving ability and student interest may result from a STEM-based inquiry curriculum, as suggested by Kellye and Knowles (2016), greater use of STEM instruction may encourage student achievement. I recommend that STEM teachers, with the support of district administrators, open up their classrooms to the community and local news organizations, to increase the public's understanding of the possible benefits of STEM training and education to student development, with no loss of mathematics achievement. I recommend that STEM instruction be adopted more widely, for the same reasons.

Because many specific STEM education teacher training programs are not locally available or affordable for teachers, I recommend that school districts provide specific STEM-based professional development training opportunities for teachers, which may increase teacher self-efficacy, which in turn may influence student achievement. Because research showed that early mathematics instruction influences mathematics achievement in secondary education, I recommend that local universities, policymakers, and other educational stakeholders invest in early-stage STEM programs in their local school districts, and support teacher in-service training programs to increase the number of STEM-confident teachers.

In addition, the results of this study could prompt school districts to examine how STEM-based mathematics instruction is delivered in various zones within the district. Since my results differed from current literature that suggested that STEM-based education, in which students apply knowledge from the classroom to real world settings through hands-on experiences, may influence mathematics achievement, it is possible that STEM pedagogy in the target district fell short of what is described in the literature. In order to ensure a rigorous and effective STEM-based mathematics curriculum, each school district that offers STEM-based mathematics instruction should analyze its program in order to meet the needs of its early learner population. Given that the target district's content standards are delivered using Polya's (1957) problem-solving heuristics, which are a core component of STEM-based tenets, I recommend that it consider examining the information resulting from this study to possibly develop interest-based curricula for populations of students who are underrepresented in STEM-based programs and STEM fields because they are relegated to traditional educational environments for one reason or another.

The findings of this study offer implications for positive social change. First, the lack of statistically significant differences in average mathematics test scores between STEM and non-STEM schools presents the opportunity for the target school district to pilot STEM-based mathematics instructional strategies to all third grade students in its

159 elementary schools. This action by the school district could have positive social change implications since prior research shows that STEM-based instructional strategies enhance student problem-solving abilities and critical functional skills, which are essential academic habits that are necessary in later academic years. Second, the findings from this study may inspire changes to the traditional mathematics curriculum, to include a more student-centered focus, and concentrate, as STEM education does, on developing student metacognitive abilities, persistence in solving challenging problems, critical thinking ability, collaborative learning, and student enjoyment (Allen-Lyall, 2018; Gravemeijer et al., 2017; Polya, 1957; Tanenbaum, 2016).

Third, the results could provide information on how different populations of students learn and how outside factors may affect their learning. Most of the target district's schools are Title I, and adopting STEM curriculum for at-risk students could result in positive social change, since research shows that students learn best when they have hands-on opportunities such as those found in STEM approaches to learning. Fourth, the focus of RQ1 was on individual changes in growth over time to determine the effects of the approaches to learning of STEM and non-STEM schools and how they influenced mathematics assessments. Given the results of no group difference in average test scores, further research is required to determine the broader effects of additional predictors such as gender, ethnicity, and socioeconomic status, or Title I designation. The data for these variables are publicly available and the results of further studies may influence positive social change in STEM education. Other variables that deserve further examination to determine prediction of state-mandated standardized mathematics test scores include income and educational levels of parents, amount of homework prescribed, classroom size, or even physical activity, or accessibility to music and art classes. Finally, the results of this study can benefit officials involved with devising interventions for populations of students who tend to score below meets standards level. Given that there were interindividual differences in test scores over time between gender and ethnicity, this study could influence positive social change by closing the achievement gap in mathematics test scores.

Conclusion

Through this study, I found that average mathematics test scores of third grade students who were enrolled in STEM schools in one urban school district in the southeastern United States were no different from mathematics test scores of third grade students who were enrolled in non-STEM schools in the same district. The empirical growth plots illustrated results from Level 1 in which each school's individual growth trajectory demonstrated fluctuations in outcomes values over time, including the variations in growth rate. Results from RQ2 revealed that there were interindividual differences and variability in average test scores between students within each school, which indicates that further research needs to be performed to determine what kinds of additional predictors or factors could be influencing the individual growth trajectory over time of each school. The predictors could be due to population differences or school-based factors. Because the Level 2 model describes the relationship between interindividual differences in the Level 1 individual growth parameters and the time-invariant characteristics of the individual, further research is warranted. Mathematics achievement of STEM students was equal to that of non-STEM students. The conclusions of this study support development of STEM-for-all programs, backed by strong teacher training in STEM pedagogy, given that STEM instruction has potential to deliver achievement similar to non-STEM instruction while inspiring the next generation of STEM-field workers through hands-on, project based early learning.

References

Abdullah, N., Halim, L., & Zakaria, E. (2014). VStops: A thinking strategy and visual representation approach in mathematical word problem solving toward enhancing STEM literacy. *Eurasia Journal of Mathematics, Science & Technology Education, 10*(3). https://doi.org/10.12973/eurasia.2014.1073a

Alden, E., Schwartz, B. L., & Strauss, R. (2016). Remedial education: Federal education policy. *Council on Foreign Relations*. Retrieved from http://www.cfr.org/unitedstates/remedial-education-federal-education-policy/p30141

Allen-Lyall, B. (2018). Helping students to automatize multiplication facts: A pilot study. *International Electronic Journal of Elementary Education, 10*(4), 391-396. Retrieved from http://files.eric.ed.gov/fulltext/EJ1176556.pdf

Anderman, E. M. (2010). Reflections on Wittrock's generative model of learning: A motivation perspective. *Educational Psychologist, 45*(1), 55-60. Retrieved from http://elizabethrodriguez.pbworks.com/f/Motivational_Article_5-1.pdf

Arik, M., & Geho, P. (2017). Skill gaps, human capital formation, and strategic choices: What do businesses do when facing critical labor force challenges? *Global Journal of Management and Marketing, 1*(2), 20. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA533556051&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=2574044X&p=AONE&sw=w

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science, 17*(5), 775-785. https://doi.org/10.1111/desc.12155

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education, 72*, 145-157. https://doi.org/10.1016/j.compedu.2013.10.020

Boaler, J. (2015). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. San Francisco, CA: Jossey-Bass.

Booth, J. L., McGinn, K. M., Barbieri, C., Begolli, K. N., Chang, B., Miller-Cotto, D., Young, L. K., & Davenport, J. L. (2017). Evidence for cognitive science principles that impact learning in mathematics. In D. C. Geary, D. B. Berch, R. J. Oschsendorf, & K. M. Koepke, (Eds.) *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 297-325). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-805086-6.00013-8

Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research – A systematic review of recent trends. *Computers & Education*, 114, 255–273. https://doi.org/10.1016/j.compedu.2017.07.004

Bryson, J. R., Mulhall, R. A., Lowe, N., & Stern, J. (2018). Engineering and the skills crisis in the UK and USA: A comparative analysis of employer-engaged education. In Y. Zhang, & M. Gregory (eds), *Value creation through engineering excellence* (pp. 327-349). New York, NY: Palgrave Macmillan.

Burchinal, M. R., Nelson, L., & Poe, M. (2006). IV. Growth curve analysis: An introduction to various methods for analyzing longitudinal data. *Monographs of the Society for Research in Child Development*, *71*(3), 65-87. https://doi.org/10.1111/j.1540-5834.2006.00405.x

Carnevale, A. P., Cheah, B., Ridley, N., & Strohl, J. (2018). *Good jobs that pay without a BA.* Washington, DC: Georgetown Center on Education and the Workforce. Retrieved from https://goodjobsdata.org/wp-content/uploads/Good-Jobs-wo-BA.pdf

Carnevale, A. P., Smith, N., Gulish, A., & Hanson, A. R. (2015). *Iowa: Education and workforce trends through 2025.* Washington, DC: Georgetown University Center on Education and the Workforce. Retrieved from http://publications.iowa.gov/20875/1/Iowa_Wrkfrce2025.pdf

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, *7*(11), E537–E540. http://doi.org/10.3978/j.issn.2072-1439.2015.10.63

Castronova, J. A. (2002). Discovery learning for the 21st century: What is it and how does it compare to traditional learning in effectiveness in the 21st century. *Action Research Exchange, 1*(1), 1-12. Retrieved from https://pdfs.semanticscholar.org/c373/76908a577387c9282673da7e2cdac87b2747.pdf

Cerda, C. A., Im, M. H., & Hughes, J. N. (2014). Learning-related skills and academic achievement in academically at-risk first graders. *Journal of Applied Developmental Psychology*, *35*(5), 433-443. https://doi.org/10.1016/j.appdev.2014.08.001

Chang, C. C., & Silalahi, S. M. (2017). A review and content analysis of mathematics textbooks in educational research. *Problems of Education in the 21st Century*, *75*(3). http://oaji.net/articles/2017/457-1498500995.pdf

Chiu, A., Price, C. A., & Ovrahim, E. (2015). *Supporting elementary and middle school STEM education at the whole school level: A review of the literature.* Paper presented at NARST 2015 Annual Conference, April 11–14, 2015, Chicago, IL. Retrieved from https://www.msichicago.org/fileadmin/assets/educators/science_leadership_initiative/SLI_Lit_Review.pdf

Clements, D. H., & Sarama, J. (2016). Math, science, and technology in the early grades. *Future of Children*, *26*(2), 75-94. Retrieved from https://files.eric.ed.gov/fulltext/EJ1118544.pdf

Confrey, J., & Maloney, A. (2015). Engineering [for] effectiveness in mathematics education: Intervention at the instructional core in an era of common core standards. In J. A. Middleton, J. Cai, & S. Hwang (Eds.), *Large-Scale Studies in Mathematics Education* (pp. 373-403). New York, NY: Springer.

Cook, B. G., & Cook, L. (2008). Nonexperimental quantitative research and its role in guiding instruction. *Intervention in School and Clinic, 44*(2), 98-104. https://doi.org/10.1177/1053451208321565

Corlu, M. S. (2013). Insights into STEM education praxis: An assessment scheme for course syllabi. *Educational Sciences: Theory & Practice, 13*(4), 2477–2485. https://doi.org/10.12738/estp.2013.4.1903

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Daily, S. (2014). *Initiatives from preschool to third grade: A policymaker's guide*. Denver, CO: Education Commission of the States. Retrieved from http://www.ecs.org/docs/early-learning-primer.pdf

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices*. Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781452272047

Darling-Hammond, L. (2017). Developing and measuring higher order skills: Models for state performance assessment systems. Retrieved from https://www.learningpolicyinstitute.org/sites/default/files/productfiles/Models_State_Performance_Assessment_Systems_REPORT.pdf

Daugherty, M. K., Carter, V., Swagerty, L., & Daugherty, M. K. (2016). Elementary STEM education: The future for technology and engineering education? *Journal of STEM Teacher Education, 49*(1), 45–55. Retrieved from http://ir.library.illinoisstate.edu/jste

Davies, S., Janus, M., Duku, E., & Gaskin, A. (2016). Using the Early Development Instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly*, *35,* 63-75. https://doi.org/10.1016/j.ecresq.2015.10.002

Davis, J. D., Choppin, J., Drake, C., Roth McDuffie, A., & Carson, C. (2018). US middle school mathematics teachers' perceptions of the standards for mathematical practice by textbook type. *International Journal of Research in Education and Science, 4*(1), 55-68. Retrieved from https://pdfs.semanticscholar.org/c238/aab571d7b0893a60eb0a511901f9d81cc4a5.pdf

Denson, C., Austin-Stallworth, C., Hailey, C., & Householder, D. (2015). Benefits of informal learning environments: A focused examination of STEM-based program environments. *Journal of STEM Education: Innovations and Research*, *16*(1), 11-15.

Dewey, J. (1915). *Schools of to-morrow*. New York, NY: E. P. Dutton.

Dockrell, J., Llaurado, A., Hurry, J., Cowan, R., Flouri, E., & Dawson, A. (2017). Review of assessment measures in the early years. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Review_of_assessment_measures_in_the_early_years.pdf

Doerschuk, P., Bahrim, C., Daniel, J., Kruger, J., Mann, J., & Martin, C. (2016). Closing the gaps and filling the STEM pipeline: A multidisciplinary approach. *Journal of*

Science Education and Technology, *25*(4), 682-695. https://doi.org/10.1007/s10956-016-9622-8

Dolan, E. L., & Collins, J. P. (2015). We must teach more effectively: here are four ways to get started. *Molecular biology of the cell*, *26*(12), 2151-2155. http://dx.doi.org/10.1091/mbc.e13-11-0675

Donohue M. C., Garnst A.C., Edland S.D., & Donohue M. (2013). Package “longpower.” *Biometrics, 53,* 937-947. Retrieved from https://cran.r-project.org/web/packages/longpower/longpower.pdf

Ejiwale, J. A. (2013). Barriers to successful implementation of stem education. *Journal of Education and Learning*, *7*(2), 63–74. https://doi.org/10.1007/978-3-319-24436-5_20

Emeagwali, N. S. (2015). Investing in STEM education: The key to economic vitality. *Techniques: Connecting Education & Careers*, *90*(3), 24-26.

English, L. D. (2017). Advancing elementary and middle school STEM education. *International Journal of Science and Mathematics Education*, *15*(1), 5-24. Retrieved from https://www.researchgate.net/publication/314489807_Advancing_Elementary_and_Middle_School_STEM_Education

English, L. D., & King, D. T. (2015). STEM learning through engineering design: Fourth-grade students’ investigations in aerospace. *International Journal of STEM Education*, *2*(1), 14. https://doi.org/10.1186/s40594-015-0027-7

Fan, S. C., & Yu, K. C. (2017). How an integrative STEM curriculum can benefit students in engineering design practices. *International Journal of Technology and Design Education*, *27*(1), 107-129. https://doi.org/10.1007/s10798-015-9328-x

Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, and future.* Washington, DC: U.S. Bureau of Labor Statistics. Retrieved from: https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematicsstem-occupations-past-present-and-future/pdf/science-technology-engineering-andmathematics-stem-occupations-past-present-and-future.pdf

Fielding-Wells, J., Dole, S., & Makar, K. (2014). Inquiry pedagogy to promote emerging proportional reasoning in primary students. *Mathematics Education Research Journal*, *26*(1), 47-77. https://doi.org/10.1007/s13394-013-0111-6

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Freire, P. (1996). *Pedagogy of the oppressed (revised).* New York, NY: Continuum.

Fullan, M., & Langworthy, M. (2014). *A rich seam. How new pedagogies find deep learning*. Boston, MA: Pearson. Retrieved from http://www.michaelfullan.ca/wp-content/uploads/2014/01/3897.Rich_Seam_web.pdf

Gravemeijer, K., Stephan, M., Julie, C., Lin, F. L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education*, *15*(1), 105-123. https://doi.org/10.1007/s10763-017-9814-6

Griffin, C. C., & Jitendra, A. K. (2009). Word problem-solving instruction in inclusive third-grade mathematics classrooms. *Journal of Educational Research, 102*(3), 187-202. https://doi.org/10.3200/JOER.102.3.187-202

Grilli, L., Panzera, A., & Rampichini, C. (2018). Clustering upper level units in multilevel models for ordinal data. In F. Mola, C. Conversano, & M. Vichi (Eds.) *Classification, (big) data analysis and statistical learning* (pp. 137-144). New York, NY: Springer.

Grinis, I. (2017). *The STEM requirements of "non-STEM" jobs: Evidence from UK online vacancy postings and implications for skills and knowledge shortages*. London, United Kingdom: London School of Economics and Political Science. http://dx.doi.org/10.2139/ssrn.2864225

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development*, *19*(1), 21-46. Retrieved from https://sites.temple.edu/cognitionlearning/files/2018/01/Gunderson-et-al-2018.pdf

Haridza, R., & Irving, K. E. (2017). Developing critical thinking of middle school students using problem based learning 4 core areas (PBL4C) model. *Journal of Physics*: *Conference Series*, 812. https://doi.org/10.1088/1742-6596/812/1/012081

Harris, B., Petersen, D., & Wulsin, C. S. (2016). *Integrating mathematical thinking into family engagement programs* (No. 1543ca9e092e49818b5f4b095463f3f0). Mathematica Policy Research.

Hausman, A., & Johnston, W. J. (2014). The role of innovation in driving the economy: Lessons from the global financial crisis. *Journal of Business Research*, *67*(1), 2720-2726. https://doi.org/10.1016/j.jbusres.2013.03.021

Heck, R. H., Thomas, S. L., & Tabata, L. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.

Hedges, L. V., & Rhoads, C. (2010). *Statistical power analysis in education research. NCSER 2010-3006*. Washington, DC: National Center for Special Education Research. Retrieved from https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf

Hefty, L. J. (2015). Applying mathematics during engineering design challenges can help children develop critical thinking, problem solving, and communication skills. *Teaching Children Mathematics*, *21*(7), 422-429. https://doi.org/10.5951/teacchilmath.21.7.0422

Henson, R. K., & Hinerman, K. (2013). *Educational statistics for longitudinal research*. London, United Kingdom: Oxford University Press.

Holzer, H. J. (2017). The role of skills and jobs in transforming communities. *Cityscape*, *19*(1), 171. Retrieved from https://penniur.upenn.edu/uploads/media/Holzer_PennIUR-Philly_Fed_working_paper_091416_%281%29.pdf

Honey, M., Pearson, G., & Schweingruber, H. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Washington, DC: National Academies Press. https://doi.org/10.17226/18612

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York, NY: Appleton-Century.

Jackson, A. & Kiersz, A. (2016, December 6). The latest ranking of top countries in math, reading, and science is out-and the U.S. didn't crack the top 10. *Business Insider*. Retrieved from http://www.businessinsider.com/pisa-worldwide-ranking-of-math-science-reading-skills2016-12

Johnson, M. S., Figueroa-Colon, R., Huang, T. K., Dwyer, J. H., & Goran, M. I. (2001). Longitudinal changes in body fat in African American and Caucasian children: influence of fasting insulin and insulin sensitivity. *Journal of Clinical Endocrinology & Metabolism*, *86*(7), 3182-3187. https://doi.org/10.1210/jcem.86.7.7665

Kainz, K. (2019). Early academic gaps and Title I programming in high poverty, high minority schools. *Early Childhood Research Quarterly*, *47*, 159-168. https://doi.org/10.1016/j.ecresq.2018.08.012

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop Velez, E. (2016). *The condition of education 2016*. Washington, DC:

National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://files.eric.ed.gov/fulltext/ED565888.pdf

Kennedy, T. J., & Odell, M. R. L. (2014). Engaging students in STEM education. *Science Education International*, *25*(3), 246-258. Retrieved from https://files.eric.ed.gov/fulltext/EJ1044508.pdf

Kermani, H., & Aldemir, J. (2015). Preparing children for success: integrating science, math, and technology in early childhood classroom. *Early Child Development and Care*, *185*(9), 1504-1527. Retrieved from https://www.researchgate.net/publication/273486367_Preparing_children_for_success_integrating_science_math_and_technology_in_early_childhood_classroom

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935-937. https://doi.org/10.1126/science.1238056

Krehbiel, D., & Piper, J. (2017). Recruiting STEM students with brief summer research experiences: An opportunity for colleges and their alumni. *Journal of STEM Education*, *18*(4). Retrieved from https://www.learntechlib.org/p/181993/

Lachapelle, C. P., & Cunningham, C. M. (2014). Engineering in elementary schools. In A. Purzer, J. Strobel, & M. E. Cardella (Eds.). *Engineering in pre-college settings: Research in synthesizing research, policy, and practices* (pp. 61-88). Lafayette, IN: Purdue University Press.

Laird, N. (2004). Random effects and the linear mixed model. In *Analysis of Longitudinal and Cluster-Correlated Data* (pp. 79-95). Institute of Mathematical Statistics and American Statistical Association. Beechwood, OH and Alexandria, VA, 2004. https://projecteuclid.org/euclid.cbms/1462106081

Lameras, P., & Moumoutzis, N. (2015). Towards the gamification of inquiry-based flipped teaching of mathematics a conceptual analysis and framework. Paper presented at the *Conference on Interactive Mobile Communication Technologies and Learning,* (pp. 343-347). http://dx.doi.org/10.1109/IMCTL.2015.7359616

Lamberg, T., & Trzynadlowski, N. (2015). How STEM academy teachers conceptualize and implement STEM education. *Journal of Research in STEM Education, 1*(1), 45-58. Retrieved from https://stemeducationjournal.springeropen.com/articles/10.1186/s40594-018-0127-2

Leong, Y. H., & Janjaruporn, R. (2015). Teaching of problem solving in school mathematics classrooms. *The Proceedings of the 12th International Congress on Mathematical Education,* 645–648. Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-319-12688-3_79.pdf

Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM*, 49(6), 937-949. https://doi.org/10.1007/s11858-017-0867-3

Maab, K., & Artigue, M. (2013). Implementation of inquiry-based learning in day-to-day teaching: A synthesis. *ZDM*, 45(6), 779-795. https://doi.org/10.1007/s11858-013-0528-0

Mann, A., Rehill, J., & Kashefpakdel, E. T. (2018). *Employer engagement in education: Insights from international evidence for effective practice and future research.* London, United Kingdom: Education and Employer Research. Retrieved from https://cica.org.au/wp-content/uploads/Employer_Engagement_in_Education.pdf

Mbodila, M., & Muhandji, K. (2012). The use of ICT in education: A comparison of traditional pedagogy and emerging pedagogy enabled by ICT's. In *Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing. Las Vegas, NV: WorldComp. Retrieved from https://www.researchgate.net/profile/Munienge_Mbodila/publication/283509521_The_use_of_ICT_in_Education_a_comparison_of_traditional_pedagogy_and_emerging_pedagogy_enabled_by_ICT's/links/563c7c6f08ae45b5d288d06e.pdf

McClure, E. R., Guernsey, L., Clements, D. H., Bales, S. N., Nichols, J., Kendall-Taylor, N., & Levine, M. H. (2017). STEM starts early: Grounding science, technology, engineering, and math education in early childhood. New York, NY: Joan Ganz Cooney Center at Sesame Workshop. Retrieved from http://www.joanganzcooneycenter.org/wp-content/uploads/2017/01/jgcc_stemstartsearly_final.pdf

McDonald, C. V. (2017). STEM Education: A review of the contribution of the disciplines of science, technology, engineering and mathematics. *Science Education International*, *27*(4), 40. Retrieved from http://files.eric.ed.gov/fulltext/EJ1131146.pdf

Meltzer, L. (Ed.). (2018). *Executive function in education: From theory to practice*. New York, NY: Guilford Publications.

Moody, V. R., & DuCloux, K. K. (2015). Mathematics teaching efficacy among traditional and nontraditional elementary pre-service teachers. *European Journal of Science and Mathematics Education, 3*(2), 105-114. Retrieved from https://files.eric.ed.gov/fulltext/EJ1107837.pdf

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., & Stohlmann, M. S. (2014). A framework for quality K-12 engineering education: Research and development. *Journal of pre-college engineering education research (J-PEER)*, *4*(1), 2. https://doi.org/10.7771/2157-9288.1069.

National Academy of Sciences. (2005). Rising above the gathering storm: Energizing and employing America for a brighter economic future. Retrieved December 17, 2017, from http://www.nsf.gov/attachments/105652/public/NAS-Gathering-Storm-11463.pdf

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

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

National Council of Teachers of Mathematics. (2006). The challenge of linking research and practice. *Journal for Research in Mathematics Education, 37*, 76−86. Retrieved from https://www.jstor.org/stable/30034842

National Council of Teachers of Mathematics. (2014). *Principles to action: Ensuring mathematical success for all*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: The National Academies Press. https://doi.org/10.17226/9822

National Research Council. (2013). *The mathematical sciences in 2025.* Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=15269#toc

National Research Council. (2014). *STEM Learning is everywhere: Summary of a convocation on building learning systems*. Washington, DC: The National Academies Press.

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly, 36*, 550-560. https://doi.org/10.1016/j.ecresq.2016.02.003

Noonan, R. (2017). STEM jobs: 2017 update. *Office of the Chief Economist, Economics and Statistics Administration*, 1–16. Retrieved from http://www.esa.gov/reports/stem-jobs-2017-update

Nunes, T., Bryant, P., Evans, D., & Barros, R. (2015). Assessing quantitative reasoning in young children. *Mathematical Thinking and Learning*, *17*(2–3), 178–196. https://doi.org/10.1080/10986065.2015.1016815

Oberle, E., Schonert-Reichl, K. A., Hertzman, C., & Zumbo, B. D. (2014). Social-emotional competencies make the grade: Predicting academic success in early adolescence. *Journal of Applied Developmental Psychology*, *35*(3), 138-147. https://doi.org/10.1016/j.appdev.2014.02.004

O'Connor, P. A., Morsanyi, K., & McCormack, T. (2018). Young children's non-numerical ordering ability at the start of formal education longitudinally predicts their symbolic number skills and academic achievement in math. *Developmental Science*, *21*(5). https://doi.org/10.1111/desc.12645

Organization for Economic Cooperation and Development. (2013). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Washington, DC: Author. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Organization for Economic Cooperation and Development. (2016a). *PISA 2015 results (Volume I): Excellence and equity in education.* Paris, France: Author. http://dx.doi.org/10.1787/9789264266490-en

Organization for Economic Cooperation and Development. (2016b). *PISA 2015 results in focus.* Paris, France: Author. http://dx.doi.org/10.1787/9789264266490-en

Organization for Economic Cooperation and Development. (2016c). *Ten questions for mathematics teachers, and how PISA can help answer them*. Paris, France:

Author. http://dx.doi.org/10.1787/9789264265387-en

Organization for Economic Cooperation and Development. (2017). *Collaborative problem solving* (PISA in Focus policy note No. 78). Paris, France: Author. http://dx.doi.org/10.1787/cdae6d2e-en

Ono, S. (1966). A mathematical analysis of motor learning: A control system model. *Japanese Psychological Research*, *8*(2), 72-89. https://doi.org/10.4992/psycholres1954.8.72

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. (Trans, by G. V. Anrep). London, United Kingdom: Oxford University Press.

Petscher, Y., Kershaw, S., Koon, S., & Foorman, B. R. (2014). Testing the importance of individual growth curves in predicting performance on a high-stakes reading comprehension test in Florida (Regional Educational Laboratory report 2014-006). Washington, DC: U.S. Department of Education. Retrieved from https://files.eric.ed.gov/fulltext/ED544678.pdf

Platas, L. M., Ketterlin-Geller, L. R., & Sitabkhan, Y. (2016). Using an assessment of early mathematical knowledge and skills to inform policy and practice: Examples from the early grade mathematics assessment. *International Journal of Education in Mathematics, Science and Technology*, *4*(3), 163-173. https://doi.org/10.18404/ijemst.20881

Polly, D., & Orrill, C. (2012). CCSSM: Examining the critical areas in grades 5 and 6. *Teaching Children's Mathematics*, *18*(9), 566-573. https://doi.org/10.5951/teacchilmath.18.9.0566

Polya, G. (1957). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.

Polya, G. (1981). *Mathematical discovery: On understanding, learning, and teaching problem solving*. New York, NY: Wiley and Sons.

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A. M. (2017). An analysis of the mathematics vocabulary knowledge of third-and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences*, *57*, 22-32. https://doi.org/10.1016/j.lindif.2017.05.011

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151-160. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.1082&rep=rep1&type=pdf

Rissanen, A. J. (2014). Active and peer learning in STEM education strategy. *Science Education International*, *25*(1), 1-7. Retrieved from https://files.eric.ed.gov/fulltext/EJ1022871.pdf

Romar, G., & Matthews, K. (2015). Considerations on US government involvement in STEM education and early childhood interventions. *Journal of Science Policy & Governance*. Retrieved from https://www.bakerinstitute.org/files/3826/download/

Rosicka, C. (2016). *Translating STEM education research into practice.* Melbourne, Australia: Australian Council for Educational Research. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1010&context=professional_dev

Rothwell, J. (2013). *The hidden STEM economy.* Washington, DC: Metropolitan Policy Program, Brookings Institution. Retrieved from http://www.brookings. edu/research/reports/2013/06/10-stem-economy-rothwell.

Schanzenbach, D. W. (2014). *Does class size matter?* (Policy brief, National Education Policy Center, School of Education, University of Colorado, Boulder). Retrieved from https://nepc.colorado.edu/sites/default/files/pb_-_class_size.pdf

Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, *10*(1 & 2), 9–34. Retrieved from http://scholarworks.umt.edu/tme%5Cnhttp://scholarworks.umt.edu/tme/vol10/iss1/3%5Cnhttp://www.math.umt.edu/TMME/vol10no1and2/1-Schoenfeld_pp9_34.pdf

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education*, *196*(2), 1-38. Retrieved from http://hplengr.engr.wisc.edu/Math_Schoenfeld.pdf

Selmer, S. J., & Kale, U. (2013). Teaching mathematics through problem solving. *Innovación Educativa*, *13*(62), 45-59. Retrieved from https://dialnet.unirioja.es/descarga/articulo/4707704.pdf

Shek, D. T. L., & Ma, C. M. S. (2011). Longitudinal data analyses using linear mixed models in SPSS: Concepts, procedures and illustrations. *Scientific World Journal, 11,* 42–76. https://doi.org/10.1100/tsw.2011.2

Simmons, R. G. (2017). *Moving into adolescence: The impact of pubertal change and school context*. New York, NY: Routledge.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. London, United Kingdom: Oxford University Press.

Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., & Kibirige, J. (2017). Student attraction, persistence and retention in STEM programs: Successes and continuing challenges. *Higher Education Studies*, *7*(1), 46-59. https://doi.org/10.5539/hes.v7n1p46

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton-Century-Crofts.

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Stanley, C. (2011). *Building the Base with Common Core State Standards*. Retrieved from https://www.slideshare.net/midlawsondle/common-core-9579863.

Stipek, D. (2017). Playful math instruction in the context of standards and accountability. *YC Young Children*, *72*(3), 8. Retrieved from https://www.naeyc.org/resources/pubs/yc/jul2017/playful-math-instruction-standards

Subramanian, R., & Clark, S. (2016). The partnership of university, industry and k-12 schools to improve awareness of STEM fields. Retrieved from https://www.hofstra.edu/pdf/academics/colleges/seas/asee-fall-2016/asee-midatlantic-f2016-subramanian.pdf

Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, *3*(2), 119–120. https://doi.org/10.4300/JGME-D-11-00075.1

Sullivan, P., Clarke, D., & Clarke, B. (2009). Converting mathematics tasks to learning opportunities: An important aspect of knowledge for mathematics teaching. *Mathematics Education Research Journal*, *21*(1), 85-105. https://doi.org/10.1007/BF03217539

Tanenbaum, C. (2016). *STEM 2026: A vision for innovation in STEM education.* Washington, DC: U.S. Department of Education. Retrieved from https://innovation.ed.gov/files/2016/09/AIR-STEM2026_Report_2016.pdf

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What’s past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352-360. https://doi.org/10.3102/0013189X14553660

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112-2130. https://doi.org/10.1111/cdev.12099

Willett, J., Singer, J., & Martin, N. (1998). The design and analysis of longitudinal studies of development and psychopathy in context: statistical models and methodological considerations. *Development and Psychopathy, 10,* 395 – 426. Retrieved from https://www.semanticscholar.org/paper/The-design-and-analysis-of-longitudinal-studies-of-Willett-Singer/51c5ef15e9cd628045d7ee3617a50668780d31b3

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *Future of Children, 26*(2), 21-35. Retrieved from https://files.eric.ed.gov/fulltext/EJ1118535.pdf

Young, A. F., Powers, J. R., & Bell, S. L. (2006). Attrition in longitudinal studies: Who do you lose? *Australian and New Zealand Journal of Public Health, 30*(4), 353-361. https://doi.org/10.1111/j.1467-842X.2006.tb00849.x

Yu, C. H., & Ohlund, B. (2010). Threats to validity of research design. Retrieved from http://www.creative-wisdom.com/teaching/WBI/threat.shtml

Zembal-Saul, C., McNeill, K. L., & Hershberger, K. (2013). *What's your evidence? Engaging K-5 children in constructing explanations in science*. Boston, MA: Pearson.

Zhao, Y. (2014). *Who's afraid of the big bad dragon? Why China has the best (and worst) education system in the world*. San Francisco, CA: John Wiley & Sons.

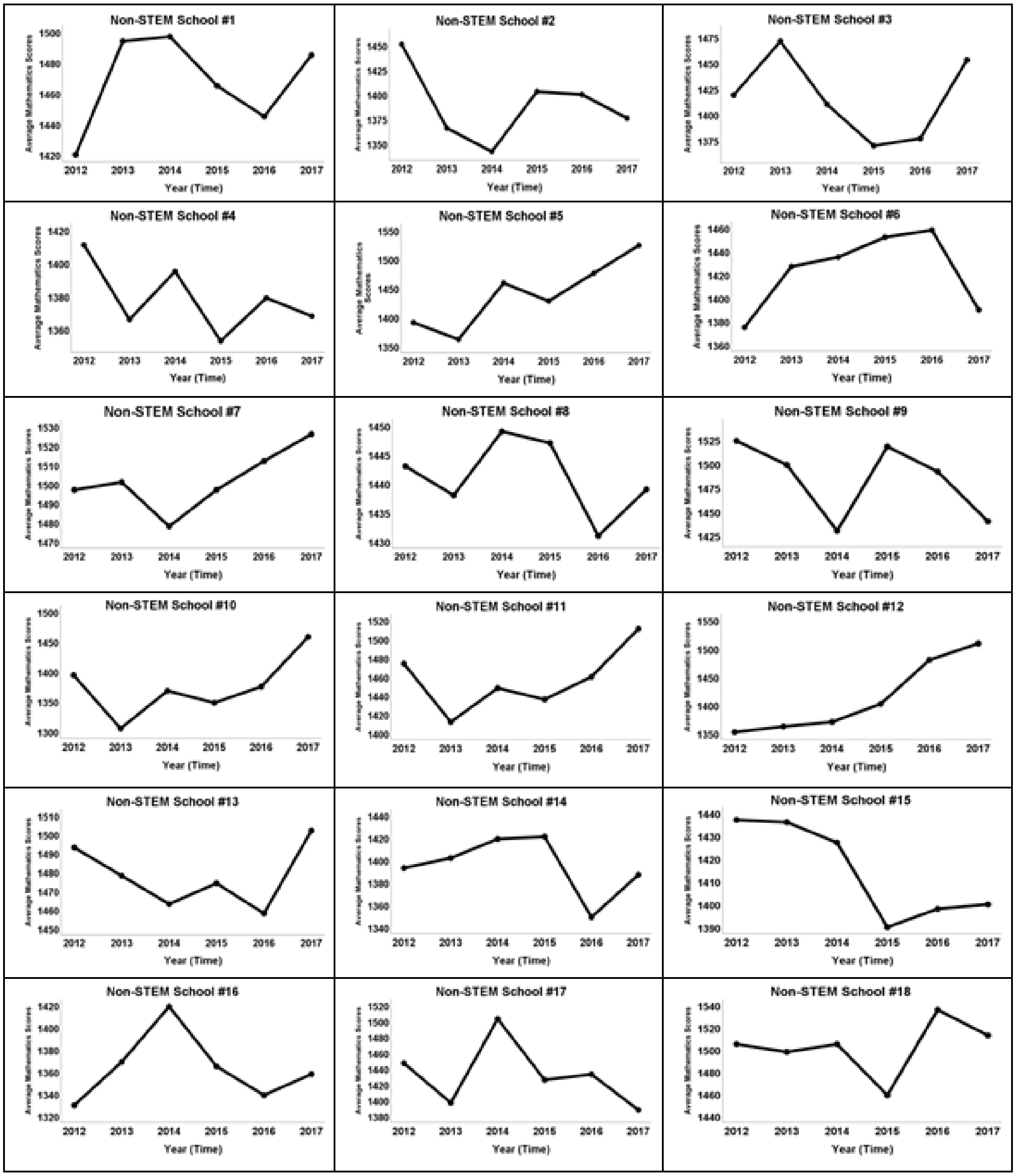
Appendix A: Statistical Program Commands for non-STEM School Analysis

Command	Syntax (non-STEM school)	Interpretation
1	mixed Math_Average_Scale_Score_NS with YearTraditional	This MIXED syntax statement will request the mixed-level analysis procedure to perform an output analysis of the math average scale score of non-STEM schools at each TIME (2012-2017).
2	/fixed intercept YearTraditional	This syntax will list the fixed-effect variables of time and school type.
3	/random intercept YearTraditional | subject(Unit_ID_Traditional) covtype(un)	This syntax will list the random-effect variable (intercept). The SUBJECT statement specifies the classification variable, the unit identification (ID, school type) and the COVTYPE statement that captures the error covariance structure type that will best fit the data.
4	/print solution testcov /method ml.	This PRINT SOLUTION syntax statement will request an output with specific results (i.e., fixed-effect estimates, its standard errors, a t-test for the parameter, and significance tests for the estimated variance components). The TESTCOV will perform significance tests for the estimated variance components. Maximum Likelihood (ML) will estimate the model.

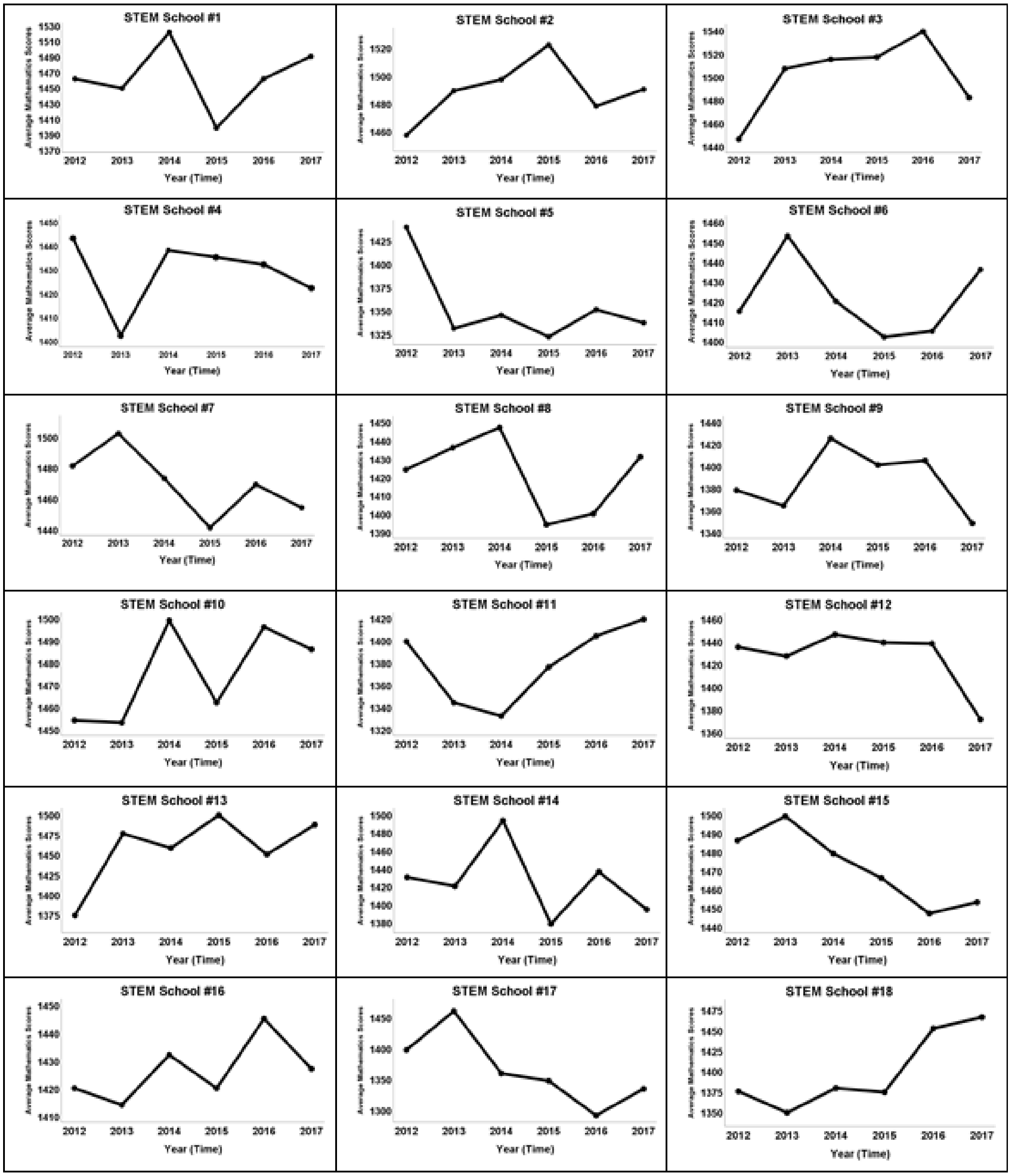
Appendix B: Statistical Program Commands for STEM School Analysis

Command	Syntax (STEM School)	Interpretation
1	mixed Math_Average_Scale_Score_S with YEARSTEM	This MIXED syntax statement will request the mixed-level analysis procedure to perform an output analysis of the math average scale score of non-STEM schools at each TIME (2012-2017).
2	/fixed intercept YEARSTEM	This syntax will list the fixed-effect variables of time and school type.
3	/random intercept YEARSTEM | subject(Unit_ID_STEM) covtype(un)	This syntax will list the random-effect variable (intercept). The SUBJECT statement specifies the classification variable, the unit identification (ID, school type) and the COVTYPE statement that captures the error covariance structure type that will best fit the data.
4	/print solution testcov /method ml.	This PRINT SOLUTION syntax statement will request an output with specific results (i.e., fixed-effect estimates, its standard errors, a t-test for the parameter, and significance tests for the estimated variance components). The TESTCOV will perform significance tests for the estimated variance components. Maximum Likelihood (ML) will estimate the model.

Appendix C: Empirical Growth Plots of Non-STEM Schools Mean Mathematics Scores



Appendix D: Empirical Growth Plots of STEM Schools Mean Mathematics Scores



Appendix E: Tables of Unconditional Mean Model of Non-STEM Schools

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
Random Effects	Intercept	1	Identity	1	Unit_ID_Non-STEM
Residual				1	
Total		2		3	

a. Dependent Variable: Math_Average_Scale_Score_NS.

Information Criteria[a]

-2 Log Likelihood	1119.890
Akaike's Information Criterion (AIC)	1125.890
Hurvich and Tsai's Criterion (AICC)	1126.121
Bozdogan's Criterion (CAIC)	1136.937
Schwarz's Bayesian Criterion (BIC)	1133.937

a. Dependent Variable: Math_Average_Scale_Score_NS.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.0	20950.2	.000

a. Dependent Variable: Math_Average_Scale_Score_NS.

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	1431.8	9.8	18.0	144.7	.000	1411.0	1452.5

a. Dependent Variable: Math_Average_Scale_Score_NS.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		1319.0	196.6	6.7	.000	984.8	1766.6
Intercept [subject = Unit_ID_ Traditional]	Variance	1541.5	588.0	2.6	.009	729.8	3255.7

a. Dependent Variable: Math_Average_Scale_Score_NS.

Appendix F: Mixed Model Analysis of Non-STEM Schools

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	YearTraditional	1		1	
Random Effects	Intercept + YearTraditional[b]	2	Unstructured	3	Unit_ID_ Traditional
Residual				1	
Total		4		6	

a. Dependent Variable: Math_Average_Scale_Score_NS.

Information Criteria[a]

-2 Log Likelihood	1111.2
Akaike's Information Criterion (AIC)	1123.2
Hurvich and Tsai's Criterion (AICC)	1124.0
Bozdogan's Criterion (CAIC)	1145.3
Schwarz's Bayesian Criterion (BIC)	1139.3

a. Dependent Variable: Math_Average_Scale_Score_NS.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.0	.630	.438
YearTraditional	1	18.0	1.1	.308

a. Dependent Variable: Math_Average_Scale_Score_NS.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	-4442.2	5595.7	18.0	-.794	.438	-16197.9	7313.4
YearTraditional	2.9	2.7	18.0	1.0	.308	-2.9	8.7

a. Dependent Variable: Math_Average_Scale_Score_NS.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		1003.934	167.320	6.0	.000	724.1	1391.7
Intercept + YearTraditional [subject = Unit_ID_Traditional]	UN (1,1)	330806788.0	191780731.1	1.7	.085	106194471.4	1030497440.1
	UN (2,1)	-164221.2	95202.9	-1.7	.085	-350815.6	22373.0
	UN (2,2)	81.524	47.2	1.7	.085	26.1	253.9

a. Dependent Variable: Math_Average_Scale_Score_NS.

Appendix G: Mixed Model Analysis of STEM Schools

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	YEARSTEM	1		1	
Random Effects	Intercept + YEARSTEM[b]	2	Unstructured	3	Unit_ID_STEM
Residual				1	
Total		4		6	

a. Dependent Variable: Math_Average_Scale_Score_S.

Information Criteria[a]

-2 Log Likelihood	1095.1
Akaike's Information Criterion (AIC)	1107.1
Hurvich and Tsai's Criterion (AICC)	1107.9
Bozdogan's Criterion (CAIC)	1129.2
Schwarz's Bayesian Criterion (BIC)	1123.2

a. Dependent Variable: Math_Average_Scale_Score_S.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.0	.190	.668
YEARSTEM	1	18.0	.022	.885

a. Dependent Variable: Math_Average_Scale_Score_S.

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	2160.2	4961.1	18.0	.435	.668	-8262.6	12583.0
YEARSTEM	-.361	2.4	18.0	-.147	.885	-5.5	4.8

a. Dependent Variable: Math_Average_Scale_Score_S.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		878.345	146.3	6.0	.000	633.5	1217.6
Intercept + YEARSTEM	UN (1,1)	239355580.7	151504420.2	1.5	.114	69224564.1	827612203.0
[subject =	UN (2,1)	-118955.9	75252.3	-1.5	.114	-266447.8	28535.8
Unit_ID_STEM]	UN (2,2)	59.1	37.3	1.5	.114	17.1	204.1

a. Dependent Variable: Math_Average_Scale_Score_S