STUDYING THE RELATIONSHIP BETWEEN TEACHERS' FIDELITY OF CURRICULUM IMPLEMENTATION AND PRESCHOOL STUDENTS' SCIENCE OUTCOMES

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Raina Kim

B.B.A. James Madison University

M. Ed. George Mason University

June, 2020

ABSTRACT

High quality preschool programs are expected to support children's learning and development. Preschool teachers have a central role in providing experiences that align with those expectations, which include teaching science to preschool children. Academic curricula in subjects such as science are useful tools that can help diffuse theory- and research-based academic instruction to preschool classrooms across geographic and socioeconomic boundaries. Teachers can use curricula in a number of ways – keep curricula as they are, adapt them to meet other goals, or as a guiding resource (Remillard, 1999). Measuring and understanding the extent to which teachers implement curricula with fidelity is necessary in order to know if the curriculum was implemented as intended and to link the teachers' enactment of the curriculum to children's outcomes (Darrow, 2013).

This dissertation focused on teachers' fidelity of curriculum implementation and how fidelity components relate to children's learning outcomes. I examined the relationship between teachers' (n = 31) fidelity to the *MyTeachingPartner-Science* (*MTP-S*) curriculum and changes in preschool children's science learning outcomes (n = 328), as well as the relationship among the fidelity components (i.e., adherence, dosage, and quality of delivery) in relation to changes in children's outcomes. Through mixed model regression analysis, I found that quality of delivery, one teacher-child interaction domain, *Instructional Support,* had a significant main effect. Adherence and dosage did not have significant main effects. In addition, I found that quality of delivery significantly moderated the associations between adherence and dosage and children's

outcomes. My findings highlight the importance of understanding how the components of curriculum implementation fidelity contribute to preschool children's science learning outcomes.

Curriculum, Instruction, and Special Education
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, "Studying the Relationship between Teachers' Fidelity of Curriculum Implementation and Preschool Students' Science Outcomes", has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Name of Co-Chair (insert name)

____
Name of Co-Chair (Jennifer Chiu)

____
Committee Member Name (Jamie DeCoster)

____
Committee Member Name (Stephanie Moore)

____
Committee Member Name (Jessica Whittaker)

________________________Date

DEDICATION

To my amazing grandmothers, mother, and aunt – who inspire me with their thirst for learning and love for others.

ACKNOWLEDGEMENTS

I would like to thank all those whose words and actions have given me the impetus and courage to begin and continue this journey. Soon Keum Kim, my mother, usually disagrees with me but has always been my biggest champion; I thank you from the bottom of my heart. Hong, Min, Yung, Hanh, Jeanne, Chloe, Calvin, Shaelei, and Jacinda, I am so grateful I have you in my life to share life's joys and struggles.

My deepened understanding of research, curriculum, teaching, and children has been made possible through the hard work of my dissertation committee members and the many *MTP-M/S* research team members. I recognize and deeply appreciate your role in my growth as a researcher, teacher, instructional designer, and an informed advocate of high-quality interactions. A special thank-you goes out to Mable Kinzie, Jessica Whittaker, Ginny Vitello, and Stephanie Moore. My sincere gratitude goes to Jennie Chiu and Jamie Decoster for your kind words, constructive guidance, and countless hours that have help me persist through these many, many years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Overview

Many U.S. states have adopted science standards for preschools, and the use of academic curricula in preschool has become more widespread. Some educators and researchers remain skeptical about whether teacher-led instruction in preschool can actually produce measurable benefits in science learning outcomes, and at the same time contribute to high-quality preschool experiences for children (e.g., Bonawitz et al., 2011; Platz & Arellano, 2011). This skepticism, coupled with mixed evidence from preschool curriculum implementation studies on fidelity of implementation and its relationship to children's learning outcomes, casts doubt on the utility of preschool science curriculum.

A necessary consideration in studies on effects of curricula on children's learning is how teachers will implement the curricula—whether teachers can use them as intended and if their *fidelity* to, or faithful implementation of, curricula's core components leads to children's learning. However, the literature has very little agreement and clarity on what constitutes fidelity of implementation of preschool curricula (Darrow, 2013, O'Donnell, 2008; Odom et al., 2010).

In order for science curricula to have their intended value, the teachers' high fidelity to researchers' and curriculum designers' ideation of expert instructional practice should help children learn science. Teachers' use of curricula has implications for teachers' complex classroom practices and thus fidelity's relationship to children's science learning can also be theorized to be complex. In this dissertation, I investigate the association between fidelity of curriculum implementation and children's science outcomes, as well as the associations among the fidelity components on children's science outcomes.

In researchers' and curriculum designers' theory of change, curricula are the necessary conduits of expert teaching knowledge and practices that result in children's learning. On their own, the expert knowledge and skills represented in the curricula are idealized interventions. Teachers are the agents who deliver the instructional experiences to children. When teachers implement the curricula as designed in the context of their classrooms, researchers and designers hope that teachers can approximate intended fidelity (Darrow, 2013; Jenkins et al., 2018; Nelson et al., 2012). The enacted curricula in the context of real classroom (Century & Casata, 2016; Pinar, 2011) should be monitored in order to determine to what extent the intended curricula came to fruition. Many preschool curriculum intervention studies have reported fidelity of curriculum implementation and a few have examined whether curriculum may have changed teachers' pedagogy. However, the construct of teachers' fidelity to preschool curriculum has remained mostly amorphous, without shared, context-specific understanding (Century & Cassata, 2016). As my literature review in Chapter 2 will show, fidelity of curriculum implementation's relation to children's academic outcomes has received underwhelming theoretical and empirical attention from the preschool curricula research community.

Need for Early Science Education

The need for science education in early childhood has come into focus in the United States in light of a growing deficit in science career professionals and a need for a science-savvy public. While science-related jobs tripled from 2000 to 2010 and projected growth rate of these jobs is on the upward trajectory, companies voice concerns that those jobs cannot be filled with qualified workers (Langdon et al., 2010). The National Science Board (NSB, 2010; NSB, 2015) of the National Science Foundation has reported the higher growth in STEM jobs than others types of jobs, as well as the growing level of science knowledge necessary even in traditionally non-science related positions (e.g., journalists, technical writers, and business analysts). Citing the high potential for growth in the science and technology sector, the United States government has long sought to maintain the country's dominant economic position by producing experts who can innovate and a general public capable of comprehending increasingly complex science- and technology-related issues (e.g., global warming, genetically modified organisms, and cyber security). The NSB (2015) reported that securing the future economic success of the country depends on making quality science, technology, engineering, and mathematics (STEM) education available to diverse populations.

One of the strategies for capacity building is supporting students in their pursuit of studies and jobs in science, technology, and engineering (Rising Above the Gathering Storm Committee, 2010). However, multi-national studies on math and science show that education in the United States has not yielded students who have globally competitive competencies in science. The results of two multi-national studies indicate that the U.S. education systems' efforts at producing a population prepared for science-related careers is not completely failing (i.e., U.S. students are not in the worst scoring tier), but that those efforts have not produced students who can be described as excelling compared to students in other wealthy countries. The *Programme for International Student Assessment* (*PISA,* Organisation for Economic Co-operation and Development, 2016), a large multinational study of 15-year-olds by Organisation for Economic Co-operation and Development (OECD) member high-income countries showed

U.S. students' average science score was not measurably different than the average score of all participating countries. The average scores of 18 countries were significantly higher than that of U.S. students. According to the 2015 *Trends in International Math and Science Study* (*TIMSS*), the largest multi-national comparative study available (Lang, 2007), fourth and eighth-grade students in the U.S. scored higher than the average amongst the 55 fourth-grade and 44 eighth-grade participating countries and other education systems (e.g., subnational entities such as Hong Kong and non-national entities such as Northern Ireland) but ranked at the bottom of ten highest scoring countries. A closer look at the science scores for U.S. students reveals that only 16% of fourth-grade students and 12% of eight-grade students performed at the *Advanced Benchmark* compared to students in top-performing Singapore, where 37% of fourth-grade students and 42% of eighth-grade students achieved the advanced benchmark (Provasnik et al., 2016). *TIMSS* compares student performance on the international benchmarks of advanced, high, intermediate, and low, with cutpoints selected near the standard 90^{th}, 75^{th}, 50^{th} and 25^{th} percentiles (Provasnik et al., 2016).

Science achievement gaps begin early and children can benefit from high-quality early science experiences. A recent study that examined children's science achievement from kindergarten through eighth grade found that a majority of children who entered kindergarten with low level of knowledge of the physical, biological, and social sciences continued to have low level of science knowledge in grades three through eight (Morgan et al., 2016). Proximal to preschool education, children's knowledge at kindergarten entry measured in the fall significantly predicted the score in the spring of first grade, and science achievement gaps that existed at kindergarten entry persisted for the majority of those students through the eighth grade

(Morgan et al., 2016). In addition, research on early childhood science education indicates positive effects of early science instruction on future science achievement for children (Tao et al., 2012).

Limited Opportunities for Children to Learn Science

However, children may have few opportunities to engage with science materials and planned science-related activities in preschool. In one study of 20 preschool classrooms, teachers intentionally provided science learning opportunities—by planning lessons, providing materials, and encouraging students to make discoveries—in less than five percent of all free-time activities (Tu, 2006). Although science materials (e.g., plants, fossils, and animals) were available in the classrooms, only about half of the classrooms had dedicated science areas with materials where children could practice science skills, such as observing, classifying, and predicting (Tu, 2006). Given that very few children at a time could engage in science exploration during free time, the combination of the lack of science stations and teachers' limited attention to children's science activities reflect an environment where science learning is not likely to occur for many children.

Studies demonstrate that children may not only have low but also uneven exposure to science. Early and colleagues (2005) analyzed data from two large datasets of state-funded preschools from eleven states during free time, small group, and whole group activities and found children engaged with science approximately 10% of the classroom time. Using one of the same datasets, La Paro and colleagues (2009) found a similarly low amount of time spent on science—an average of 7% of the observed classroom days was spent on science. Noteworthy is that La Paro and colleagues found that some students spent no time on science while others spent as high as 26% of their time, indicating uneven opportunities for children to learn science.

The federal and state governments' aim to provide students in kindergarten through high school with higher quality of science instruction and more equitable learning opportunities (e.g., National Research Council, 2013; U.S. House, Committee on Science, 2001) needs to be extended to preschools. Science curricula are tools that can help in this effort, but little research has been conducted on preschool science curricula, likely due to the interest in preschool science education being newer in comparison to language arts and mathematics. Standards and advice from government and non-governmental organizations can guide science education efforts and provide a foundation upon which science education efforts are built. In addition, design and implementation efforts of preschool curricula require an understanding of the history and prevalent philosophical beliefs about high quality preschool education.

Science Standards: What Should Preschool Children Learn?

In recent years, proponents of science education have worked to institutionalize science learning in children's academic progression through state-mandated science learning standards and advice on science teaching. One device for disseminating this agenda into classrooms is the science curricula. Early learning specialists at state departments of education indicated that the primary intent for standards is to improve instruction and curriculum (Scott-Little et al., 2007). States have made efforts to align curricula with standards, for example, in an alignment analysis of existing curricula and selection of new curricula. Use of high-quality curricula can be an effective means of improving outcomes for young children (NAEYC & NAECS/SDE, 2003). It follows then, that high quality science curricula help teachers improve their science teaching practice in ways that benefit children.

Government directives for high-quality research, and articulation of quality instructional practices for pre-kindergarten teachers, has led to setting of standards and interest in the use of curricula (U.S. House of Representatives, 2001; NAEYC,2009; Scott-Little et al., 2007). In most states in the U.S., curricular standards adopted at the preschool level include science in order to better prepare children for science learning in later grades (Stipek, 2006). Governing and advisory bodies at the federal and state levels have tried to implement science standards that translate to beneficial experiences for children in preschool classrooms (e.g., Kansas State Department of Education, 2014; Michigan State Board of Education, 2013; National Research Council [NRC], 2013; Office of Head Start, 2010). The selection and implementation of standards and curricula remain hotly debated issues in the U.S., and thus, the adoption of standards and subsequent changes to curricula are at times in flux.

Currently, there are two major sources of guidance for what and how science should be taught in the K-12 classroom: 1) the Common Core State Standards (CCSS; National Governors Association Center for Best Practice & Council for Chief State School Officers, 2010) and 2) the Next Generation Science Standards (NGSS; NRC, 2013). The rationales behind each of the two standards are complex, but one fundamental way in which the two are different is in their focus on different aspects of science education. The CCSS focus on the skills and knowledge students should learn in math and English language arts and for grades K-5. Science is integrated into reading standards without science-specific learning expectations but with experiences in critical thinking, problem-solving, and analyses related to non-fiction reading (NRC, 2013). In contrast, the NGSS are based on NRC's (2012) framework which incorporates the way scientists work and think into the expectations for science knowledge and skills acquisition. In this framework, students learn science through engaging in scientific practices (i.e., developing explanations, generating and evaluating evidence and explanations, and participating in the practices and discourse of science). The NGSS provide expectations for student performance (i.e., what students will know and be able to do), without curricular guidance (e.g., recommending use of reading materials). Since these two types of standards are not contradictory, they can be simultaneously addressed by curricula. By 2013, 41 states and the District of Columbia had adopted CCSS (Achieve, 2013) and by 2019, 19 states and the District of Columbia had adopted the NGSS (Achieve, 2019).

In addition, the federal Head Start programs have implemented the Early Learning Outcomes Framework which provides guidance for programming, including the selection of curricula. The framework outlines five central domains of learning and development—approaches to learning; social and emotional development; language and literacy; cognition; and perceptual, motor and physical development. Knowledge and skills around scientific inquiry (e.g., identifying five senses and using senses to describe phenomena) and scientific reasoning and problem solving (e.g., making predictions based on prior knowledge and experiences) are included in the cognitive domain (Office of Head Start, 2015).

The state content standards for kindergarten and guidance from the Office of Head Start have been used to provide guidelines at the preschool level by some states. For example, California State Department of Education's California Preschool Learning Foundations aligns with Common Core Standards, kindergarten state content standards, and the Head Start Child Development and Early Learning Framework (California Department of Education, 2008).

There is some variability, but the state preschool science standards in states such as Kansas, Massachusetts, and Michigan include topics in physical, earth, and life science, and focus on science process skills similar to skills outlined by the Office of Head Start, such as observing, describing, collecting data, exploring/testing, and inquiring (e.g., Office of Head Start, 2015; Kansas State Department of Education, 2014; Massachusetts Department of Education, 2003; Michigan State Board of Education, 2005).

Curricula can be powerful tools for dissemination of high-quality instructional practices and can contribute to equalization of science learning opportunities. Although evidence is mixed, the indications that children can benefit from teachers' use of science curricula, coupled with support of standards and advice, are justifications for designing, implementing, and improving preschool science curricula. The existing standards and framework provide a foundation for developing and designing curricula that would help preschool teachers engage children in science learning. However, curricula are simply tools. Teachers' use of curricula in teaching science to preschool children includes complex, if not challenging, phenomenon of many types of strategies and adjustments of strategies.

Preschool Teachers' Practice

In asking teachers to use curricula with fidelity, it is important to understand the context in which they work. Preschool classrooms are unlike the upper grades, where curricula on academic subjects have long been part of teachers' practice and academic learning outcomes have long been prioritized. Traditionally, preschool teachers have been responsible for creating high-quality classroom experiences for children through strategies such as communicating warmth and fostering children's independence – strategies that are sensitive to children's developmental needs (e.g., Hindman et al., 2010; Zaslow et al., 2011). However, the same ideas that have supported developmentally appropriate practices (DAP) and prioritized them have also hindered the inclusion of academic instruction, including science instruction, in preschools until more recently.

Young children observe and experience the world around them with curiosity, building knowledge about their world. Whether preschool children should be taught academic subjects such as science through teacher-led instruction or left to observe and experience the world at their own volition has been a long-standing point of contention and debate in preschool programming. What constitutes an appropriate match between learning activities and children's development varies widely depending on philosophies and theories (Katz, 2003). Although not the only influential developmental theory, Piaget's stage theory in particular informed early recommendations on DAP and shaped the prevailing understanding of what preschool children can learn and how their development should be supported in preschool. Piaget's theory on developmental stages is useful to illustrate children's potential development over time, but it also suggests a constraint as to what children are capable of learning based on age-related assumptions about cognitive capabilities (Metz, 1995).

The National Association for the Education of Young Children (NAEYC; a leading accreditation and membership organization for early childhood programs and professionals) has advised early childhood educators and administrators with guidance on DAP. Several revisions to DAP have emerged over time, reflecting the prevailing perceptions of children's capacity for various kinds of learning and the practices that support it. One of the concerns reflected in past NAEYC position statements on DAP is that academic learning (i.e., in academic subjects such language and literacy, math, and science) promoted through standard-setting and curriculum implementation may impede children's development in physical, social, emotional, and cognitive domains which are interrelated (NAEYC, 1996; NAEYC, 2009). In 1986, NAEYC published its first description of DAP to guide educators in the best pedagogies for early education (NAEYC, 1986). NAEYC, along with many educators and policy-makers, voiced concerns about primary school curricular standards being pushed down to kindergarten and preschools without regard for the developmental readiness of young children (Shepard & Smith, 1988). The 1986 version of DAP favored teacher practice in which the teacher does not teach directly but creates an environment for children to learn through exploration at the children's choosing and pace (NAEYC, 1986). Some scholars asserted that formal instruction in early education would deprive children of the opportunity to play and learn from *natural* exploration while subjecting them to passive and rote learning experiences (e.g., Shepard & Smith, 1988). Edward F. Zigler, a developmental psychologist and the first director of the U.S. Office of Child Development, voiced concerns about "depriving [young children] of their most precious commodity—their childhood," and warned against formal and structured academic instructional time (Zigler, 1987, p. 257). The inclusion of academics to serve the interest and needs of the general population of four-year-olds remained a controversial topic (e.g., Futrell, 1987; Kagan, 1989; Zigler, 1987).

One objection to academic instruction is that children should learn about the world through self-directed play and therefore teacher-led instruction is not developmentally appropriate for preschool children. The child-driven and unstructured nature of play can be at odds with intentional teaching guided by learning objectives. The discourse on play and teaching continues today with examination of methodologies for guiding play in the classroom, exploration of new types of materials (e.g., game-based learning), and evolution of venues that provide opportunities for learning through play outside of the classroom (e.g., museums) (Sutterby & Kharod, 2019).

By 1996, the NAEYC's position statement on DAP shifted to include an emphasis on social and cultural context within which children learn, as well as support of curriculum and teaching practices that align with academic standards. This shift has continued and today's prevailing conceptions of children's cognitive development include those that are domain-specific (e.g., biological) and influenced by children's experience with their environment (e.g., Duncan, 1995; Glassman & Zan, 1995; Zelazo & Frye, 1997). In its latest 32-page position statement, the NAEYC (2009) advocated that teachers should actively support multiple aspects of children's development by preparing and providing purposeful learning experiences and considering children's individual needs. Previous cautions against too much time spent on teacher-guided instructional activities have been replaced with DAP advice which clearly outlines the many responsibilities of preschool teachers for creating a positive social environment as well as actively engaging students in skill development, drawing upon academic curricula that are based on state standards. This shift in DAP advice supports the use of science curriculum in preschool classrooms.

The DAP and the science standards can both guide what and how children should learn about science. Science curricula could be designed in consideration of what and how preschool teachers should teach and propagate high-quality science education across classrooms. But although there are some indications that children could benefit from teachers' implementation of curricula, very few evidence- and theory-based science curricula have been designed, implemented, and evaluated in preschool classrooms.

Benefits of Early Science Curricula

Not many early science curricula have been developed and evaluated with learning outcomes in published studies. Although evidence is not abundant, studies indicate the potential that well-designed science curricula could have multiple benefits for children's learning outcomes in subject-specific skills such as language, science, and mathematics, as well as in foundational skills like approaches to learning (Bustamante et al., 2018; French, 2004; Greenfield et al., 2009). Studies support the theory that children benefit when teachers can guide children in learning science, engage higher-order cognitive skills, and help children apply knowledge and skills.

In an investigation of children in Head Start centers in one urban county, Bustamante et al. (2018) found a bidirectional predictive relationship between gains in science learning outcomes and gains in approaches to learning skills. Thus, when teachers purposefully scaffolded children's science learning and tended to children's domain-general skills of approaches to learning (e.g., sustained focus, planning, and verbalizing), both science domain specific learning outcomes and domain-general skills may benefit. Through a review of literature relevant to designing activities for children, Kinzie et al. (2015) concluded that activities should allow for a balance of teacher-directedness and child-centeredness in children's learning process. In addition, activities that help children focus on specific knowledge and skills and are relevant to children's life experiences are likely to support children's learning (Kinzie et al., 2015).

Teachers' implementation of high-quality science curricula can be a means through which the potential benefits of science education for children can be realized across many classrooms. A few studies on preschool science curricula collectively indicate that science curricula can benefit children's learning outcomes (e.g., French, 2004; Greenfield et al., 2009). Studies on science curricula provide evidence that students can learn in other domains through science. For example, *ScienceStart!* (French, 2004), was designed to emphasize science and language development, while incorporating other topics such as math and social studies and practice of higher order cognitive skills. Results demonstrated significant language gains during the academic year. Greenfield et al. (2009) investigated the use of direct instruction of foundational knowledge as well as guided discovery and inquiry-based exploration of science topics to help children learn science as well as increase skills in other readiness domains (Greenfield et al., 2009). In a half-year implementation of a science curriculum, children in intervention classrooms with direct instruction performed significantly better on science and other readiness outcomes than in control classrooms (Greenfield et al., 2009).

Examining how teachers' curriculum implementation benefit children's learning outcomes could help us to improve curriculum design and support teachers' faithful implementation. Research demonstrates that teachers' fidelity of curriculum implementation is an important construct to examine (e.g., Hamre et al., 2010; O'Donnell, 2008; Preschool Curriculum Evaluation Research Consortium [PCERC], 2008). Measuring fidelity of curriculum implementation can help evaluate the curricula in various ways. For example, it can help gauge whether real teachers can use the materials with real children as intended by the researchers and designers. It can also be used to show that teachers use more desirable strategies when teachers closely follow the curricula. These desirable strategies would reflect the core ingredients of curricula that researchers theorize to be important for teachers to enact in their classrooms. Teachers' fidelity to curriculum's core components are expected to positively influence children's learning gains by helping teachers use instructional strategies that benefit children's learning.

Problem Statement

Despite the advice from NGO and government which support intentional academic teaching and use of science curricula, my literature review of published studies on preschool curriculum and fidelity of implementation showed very few science curriculum studies among them, as very few preschool science curricula exist. Also, while many studies on preschool curricula have reported whether a curriculum had treatment effect, few studies have explored the components of fidelity of curriculum implementation and how they work in their relationship to children's learning outcomes.

Curricula are devices that can be examined during use and can be improved through study. Fidelity has remained a broad and flexible construct in curriculum implementation studies, often used only as a means to show whether the event that the researchers theorized to be important took place. Limited agreement or standards on what must be measured during curriculum implementation exits in the preschool curriculum research community. The lack of clarity and shared understanding of what fidelity means in the context of teachers' practice, and how fidelity components work to help children learn, has hampered our understanding of curriculum as a tool. A more detailed examination of fidelity's relationship to children's learning could inform how researchers and designers could strengthen the relationship between the practices theorized to be important for teachers to enact and children's outcomes. For example, if one aspect of fidelity of curriculum implementation had a positive effect on children's learning outcomes or if it strengthened the relationship between another fidelity component and children's learning outcomes as theorized, then researchers and designers would want to find ways to better support that one component and/or make sure teachers could implement with high levels of both fidelity components. If one fidelity component had no effect on children's learning outcomes or if it had a negative effect on children's learning outcomes contrary to their change theory, then researchers and designers would want to try to delve deeper into their change model to identify the limitations or alternate models that better represent how teachers' use of curricula help children learn. As my literature review in Chapter 2 will show, not many studies in curricular interventions have examined fidelity of curriculum implementation in preschools to inform which components are important for children's learning and in which contexts. An important next step for preschool curricular intervention studies would be to more closely examine components of fidelity, beyond reporting it, to inform if and how the curriculum's theorized and evidence-based components are related to children's learning outcomes.

Purpose of the Study

Further examination is needed to understand how fidelity components, both individually and in combination, contribute to children's knowledge and skills. Through a quantitative study, I will test my theories about how the components of fidelity work in their relationship to children's science learning gains. By examining the relationships of components of fidelity to children's learning outcomes and mechanisms of fidelity components in their relationships with children's learning outcomes, I hope that this study will contribute to a better understanding of the role of fidelity and the processes of fidelity in children's learning outcomes.

Relevance of the Study

This study contributes to a limited body of exposition and research on preschool teachers' science curriculum implementation. It is believed that through teachers' high fidelity of curriculum implementation, curricula can ultimately enrich learning experiences across all classrooms. However, as my review of literature in Chapter 2 will show, not many studies in curricular interventions have examined fidelity of curriculum implementation in preschools to inform which components are important for children's learning and in which contexts. An important next step for preschool curricular intervention studies would be to go beyond reporting fidelity levels and to closely examine components of fidelity. Through closer examination of components of fidelity, researchers can inform if and how the curriculum's theorized and evidence-based components are related to children's learning outcomes.

This study will inform how fidelity to preschool curriculum can be conceptualized in research and curriculum design and contribute to a better understanding of the phenomenon of teaching with curriculum. The ultimate benefit of this study and related studies will be the better understanding of the contribution of instruction strategies of curriculum implementation to science outcomes; this understanding will in turn enhance researchers' and designers' ability to support teachers' practice through curriculum design and professional supports. For example, indications that fidelity components have individual contributions to outcomes or that fidelity components interact in their associations with children's outcomes may help lend support to measuring those components of fidelity in studies and focusing on ways to support those components in teachers' implementation. I hope that this study will be a step toward helping researchers and designers conceptualize preschool teachers' fidelity in specified way that can close the gap between curriculum implementation studies and preschool quality research.

Conceptual Framework

Curriculum

The term *curriculum* can convey a range of meanings, from the entirety of a curricular program that spans two or more years with state content standards and associated instructional materials (e.g., books, kits, and teacher's guides), to a collection of components from different pre-packaged materials or lessons and materials designed by individual teachers for their classrooms (Beauchamp, 1982; NRC, 1999, 2004). In this study, a *curriculum* is a set of guidances in the form of sequenced learning objectives, activities, and materials provided to the teachers to use in the classroom over a semester up to one academic year. Interventions that include components beyond that, for example belonging to a multi-year series of curricula or requiring special parent-teacher interactions, will be referred to as *programs* when the distinction is necessary.

Curricula of interest are high-quality, subject-specific curricula that are designed to align with targeted learning goals such as those put forth by national or states' standards for an academic subject. These curricula leverage theoretically grounded and evidence-based content- or subject-specific instructional practices, in order to provide uniform content and teaching methodologies that promote student learning across classrooms and schools. Designing the curricula from theory and researched evidence that also align with progressions and goals set forth in standards can provide teachers with clearly organized and explained subject-matter content and the best pedagogical guidance for the topic, including sequenced learning objectives.

Fidelity

The term *fidelity* is used broadly to mean the degree to which core components of intervention that are theorized be most important are implemented in context, and as intended. Fidelity is an important construct in intervention studies targeting children more generally, in and outside of the classroom. Implementation involves a "specified set of activities designed to put into practice an activity or program of known dimensions." (Fixen et al., 2005, p.5). However, the term lacks conceptual uniformity and specificity beyond this broad definition (Gearing et al., 2011).

In implementation science literature for early childhood, Hulleman et al. (2013) and Downer (2013) made a distinction between *intervention fidelity* as specific to the practitioner's delivery to the subject (in the case of curriculum implementation, the teachers' delivery to the students) and *implementation fidelity* as concerned with the contextual or system-level support for the intervention, such as giving teachers planning time. Although this distinction has not been widely used, the distinction is a useful one to make, as curricular intervention studies can involve an array of complementary and supplemental program components that occur outside of the classroom—such as coordination with school or district administrators and efforts to familiarize teachers with the curricular materials—that could be of interest to implementation researchers. This distinction highlights the importance to being clear about *who* and *what* are the focus in fidelity studies.

Definitions of fidelity in K-12 curriculum literature include variations on the themes of extent and measure of enactment of curricula's critical components as proposed, planned, intended, and idealized by designers and developers (Century et al., 2010; O'Donnell, 2008;). As I will show in my literature review, fidelity has been defined differently by different researchers—for example, as a component of implementation or as in this study, an overarching term with component parts, and the definitions rely largely on the school-based intervention research.

In curriculum implementation, the ultimate interventionists of interest are teachers, even when during the initial stages of the research and refinement of curricula, the researchers or members of the research team may act as interventionists. Teachers use a curriculum with the students in their classrooms with the whole class or in small groups, implementing the same lesson in each small group session. Whereas an interventionist who works with certain students might focus on the same set of skills over many sessions, progressing as it suits specific students or research goals, a teacher who uses a curriculum with their students focuses on the goals of each lesson and progressing to the next lesson in the next session with the class.

My study is interested in teachers' use of curricula in the context of their classrooms. For the purpose of this study, teachers' curriculum implementation have the following conditions: 1) interventionists are preschool teachers; 2) the focus of their classroom practices is all of their students who are recipients of teachers' enactment of curriculum; and 3) curricula provides a sequenced academic content and pedagogies organized into lessons—which may overlap but are not meant to be repeated—that should be implemented over a specified time and frequency.

When teachers interpret and use the curricula through the lens of their own experience and expertise in the unique context of their classrooms full of real students, *fidelity of curriculum implementation* or *curriculum implementation fidelity* can be observed in their practice, or behavior. Fidelity of curriculum implementation refers to the faithfulness to core components of curriculum with which teachers enact curriculum in their classrooms, as intended by the authors of the curriculum (Superfine et al., 2015).

Connecting Teachers' Instructional Strategies with Curriculum Implementation

Goals of curricular researchers and designers are not to have teachers robotically follow the guidelines given in the curricula, but rather, to aid teachers in their professional practice through providing a well-thought-out, accessible, useful instructional tool that can be used by teachers.

At the preschool level, high-quality teaching employs developmentally appropriate practices that support children's socio-emotional and cognitive needs as well as helping children reach the developmental milestones set by state standards (NAEYC, 1996). Conceptualization of curriculum implementation in the context of preschool classroom practices can be derived from preschool program quality research that has examined: 1) learning opportunities within which children are exposed with some frequency to science instructional interaction and quality of those interaction (La Paro et al., 2009), and 2) teacher practices composed of *content-specific strategies*, specific to academic subject, and *generalized teaching strategies*, observable across all types of activities (Hamre & Pianta, 2007; Hamre et al., 2013).

Then, high-quality preschool science curriculum can be conceptualized to be one in which teachers use prescribed science instructional interactions with prescribed dosage, in concert with high-quality generalized teaching strategies that target developmental needs. When teachers implement high-quality preschool science curriculum as prescribed, children are exposed to higher frequency of goal-directed science instruction, delivered with high quality content-specific and generalized teaching strategies.

Implementing a high-quality academic curricula with high level of fidelity would provide the affordances of necessary content-specific strategies and some degree of affordance for developmentally appropriate generalized teaching strategies. In the curriculum with step-by-step guidance and script, science content-specific strategies may overlap with developmentally appropriate generalized teaching strategies. For example, the guidance to prepare materials including paper towels for a hands-on floating and sinking activity addresses the requirements of the science-specific activity as well as children's developmental needs. However, in order for the generalized teaching strategies to be of the highest quality, they must be planned and adjusted to children in particular classrooms. Although limited in its ability to include through prescription, strategies that require in-context accommodations to developmental needs of the specific children in teachers' classrooms, a high-quality academic curriculum can be designed so that science-content specific strategies coincide with strategies that consider the developmental needs of archetypes of preschool children and preschool classrooms (e.g. *MyTeachingPartner-Science* [*MTP-S*], Kinzie et al., 2014). During implementation, teachers can draw on the generalized teacher-child interaction strategies embedded in the curricula along with their own strategies to meet the unique combination of the developmental needs of children in their classrooms.

Fidelity Components: Adherence, Dosage, and Quality

Above, I described teachers' practices in implementation of curriculum. Next, I define the fidelity components that are often examined in preschool curriculum studies in the context of preschool teacher practices. As in other preschool curriculum implementation studies, the definitions are derived from school intervention program implementation studies. Then, I describe how the effects of the fidelity components on children's science outcomes are conceptualized in my study.

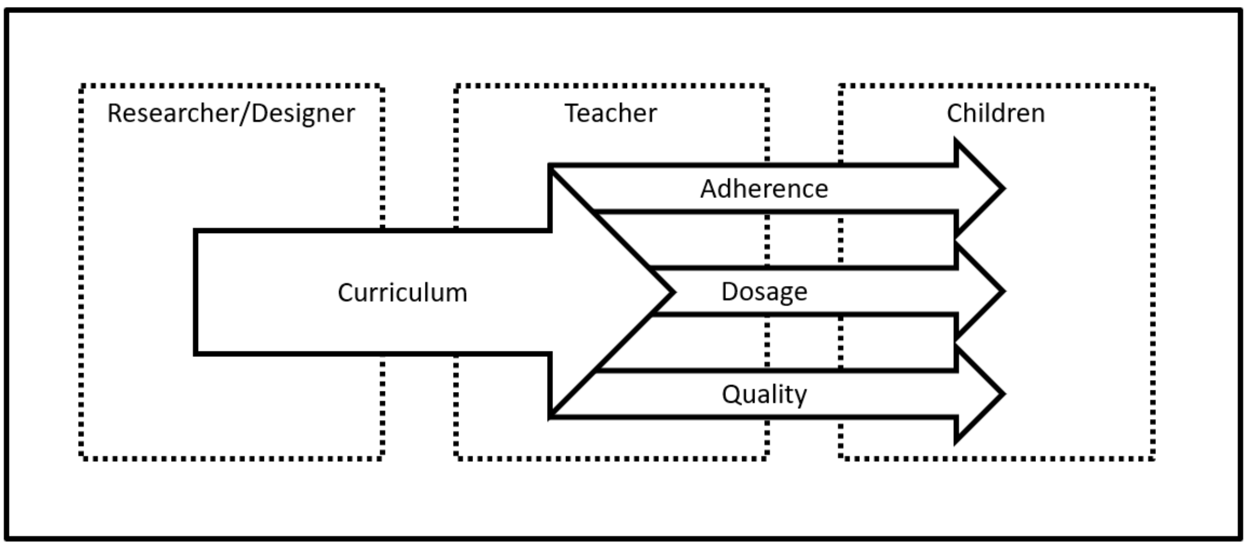
Among the possible fidelity components, I selected three that I found to be the most prevalent in preschool curriculum implementation studies for my study. In Darrow's (2013) review, although most preschool curriculum studies that measured fidelity did not examine all five components, they often measured adherence and/or exposure, and a few also measured quality of delivery. Among the 14 preschool curricula in the PCERC study (2008), adherence, dosage, and quality of delivery were the most frequently measured of five fidelity components that also included participant responsiveness and program differentiation (Darrow, 2013). I based my definitions on those that were extrapolated in O'Donnell's (2008) review and I build on those broad definitions which are applicable to interventions with specifics of curriculum implementation.

- **Adherence** is whether teachers delivered the components (e.g., structure, guidance, steps, and script) of the curriculum as prescribed;
- **Dosage** is the prescribed number, duration, or frequency of curriculum lessons teachers used with children; and
- **Quality of delivery** is the manner in which the teachers used the curriculum using generalized teaching strategies that meet children's developmental needs.

Figure 1 represents how a curriculum that originates from the researchers/designer is experienced by children. Teachers' enactment of curriculum includes components of fidelity—adherence, dosage, and quality of delivery—which are experienced by children in their classrooms.

Figure 1

Curriculum Implementation



While dosage and adherence are clearly prescribed in the curriculum (e.g., via strategies communicated through steps and scripts, and number of lessons) the quality component requires further elaboration for preschool classroom context. In preschool program research, several studies provide evidence of the role that one type of generalized teaching strategies, *teacher-child interactions*, has on children's learning outcomes (e.g., Burchinal et al., 2010; Curby et al., 2009; Hatfield et al., 2016; Nix et al., 2013; Sabol et al., 2013). Teacher-child interaction strategies can be applied across all manner of activities from direct instruction to free center time (Pianta, La Paro, & Hamre, 2008).

One measure of teacher-child interactions is the *Classroom Assessment Scoring System (CLASS*, Pianta, La Paro, & Hamre, 2008). Within *CLASS*, teacher-child interactions are described by three categories of strategies, or domains. Three major domains of teacher-child interaction are *Emotional Support*, *Classroom Organization*, and *Instructional Support* (Pianta, La Paro, & Hamre, 2008). The *Emotional Support* domain is composed of teacher-child interactions that focus on teachers' creation of warm, congenial, and emotionally expressive/sensitive classroom. The strategies in *Classroom Organization* focus on how teachers make use of time, routines, and materials and direct children's' attention and interest. The *Instructional Support* domain is composed of teacher's encouragement of deeper understanding, higher level thinking, and use of language.

There are multiple scenarios of teachers' implementation wherein the researcher's theory of change would not take place. For example, although in researchers' theory of change, fidelity components may include adherence, dosage, and quality of delivery, adherence to curriculum may not be accompanied by high quality of delivery even when the curriculum includes specific strategies that align with developmental needs such as open-ended questioning (Morris et al., 2016). Alternatively, children's science learning experiences and outcomes in classrooms with very few instances of science teaching but where teachers use high-quality teacher-child interactions would be very different from a classroom where the teachers steadfastly deliver every lesson in the curriculum and completely adhere to the guide provided in the curriculum. Whether or not these differences matter on their own or in relation to one another for children's learning is the focus of my study. Understanding individual contribution of each fidelity component and whether or not one provides the context for strengthening or weakening another component's relationship with children's learning outcomes could help guide researchers' design and support choices during curriculum development and implementation.

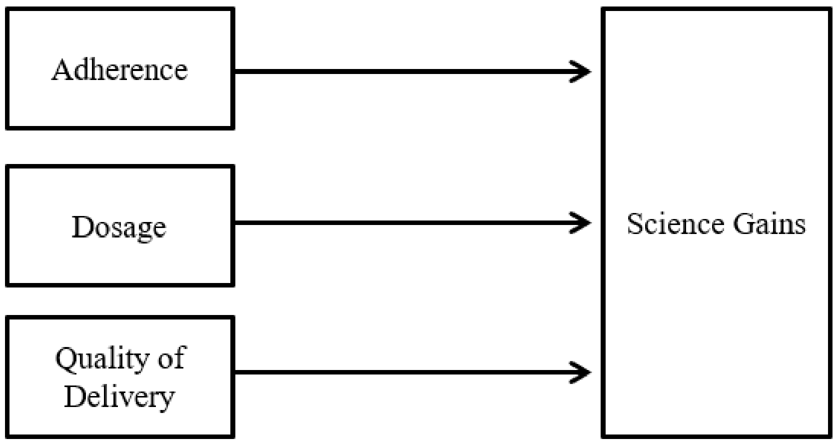
Additive Effects of Adherence, Dosage, and Quality

Implementation to research-based curriculum is intended to provide higher quality instructional experience for children (Clements & Samara, 2008; Hamre et al., 2010). The assumption in disseminating curricula is that following the curriculum closely and implementing most if not all of the lessons would result in gains in children's learning outcomes. However, in real classrooms, teachers may have competing goals that diminish fidelity of curriculum implementation and as a result diminish curriculum's utility. Little is known about which components of fidelity account for the variability in outcomes. Although theory of change in curriculum implementation assumes that teacher's high level of fidelity of implement is related to children's outcome, there are mixed indications about which fidelity components have effect on children's learning outcomes. For example, Guo et al. (2010) found that dosage (as conceptualized in the present study) was not significantly associated with children's outcomes whereas Piasta, Justice et al. (2015) found that dosage significantly predicted children's outcomes. Hamre et al. (2010) found that dosage and quality each significantly predicted children's outcomes, but adherence did not.

I posit that these frequently studied components of fidelity of curriculum implementation—adherence, dosage, and quality—each have an additive effect on children's outcomes. See Figure 2.

Figure 2

Additive Effects of Fidelity Components



Interactive Effects of Adherence, Dosage, and Quality

I consider each fidelity component to be important in order for curriculum to produce gains in children's science learning outcomes. I also posit that the effect of the prescriptive components of fidelity of curriculum implementation, adherence and dosage, will vary by teacher's ability to use high-quality teacher-child interaction strategies. In addition, I posit that the effect of teachers' adherence will be higher when teachers implement more lessons (dosage).

As teachers enact the curriculum with high adherence to the pedagogies laid out in the curriculum, children may learn more science when teachers use high-quality teacher-child interaction strategies; for example, helping children focus their attention. High adherence to the pedagogies laid out in the curriculum would have a stronger effect on children's science learning outcomes when teachers use higher dosage of curriculum, thus providing more opportunities for children to learn science. When teachers provide more science instruction from the curriculum through high dosage while using high-quality teacher-child interaction strategies to, for example, provide children with an organized and comfortable environment, children may learn more science. In this way, a given fidelity component can provide the context for maximizing another fidelity component's relationship to children's outcome.

In addition to the additive effects of fidelity components, how the relationship among the fidelity components influences children's learning gains could provide a more nuanced insight into the complexity of teachers' practices. Interpreted through the lens of teacher knowledge and practice, such moderation effects would have ramifications for researchers and designers, who may focus on science pedagogies in consideration of developmentally appropriate practice. If, for example, adherence positively predicts outcomes only in the condition of high dosage or high quality of delivery, then researchers would need to pay closer attention to how to support those components in order to maximize the curriculum's benefit to children. See Figures 3, 4, and 5 for graphic representation of the interactive associations.

Figure 3

Interactive Effect of Adherence and Quality of Delivery

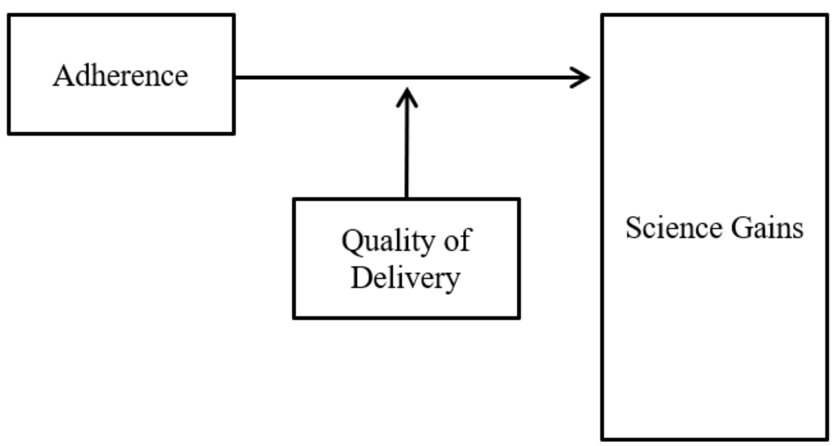


Figure 4

Interactive Effect of Dosage and Quality of Delivery

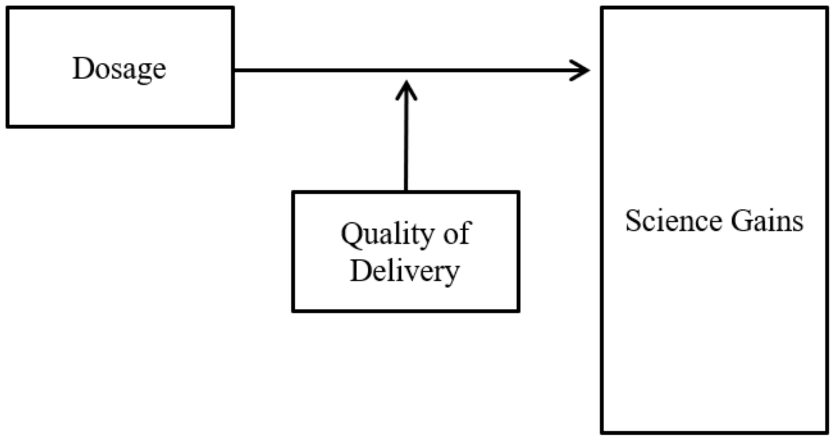
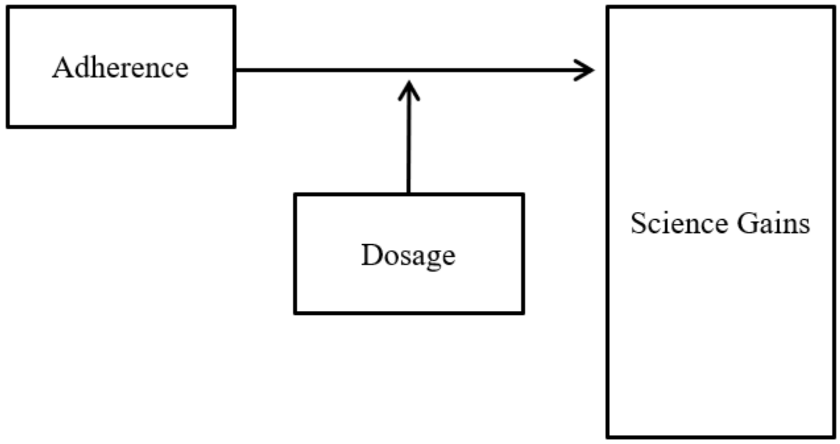


Figure 5

Interactive Effect of Adherence and Dosage



Research Questions

Based on the framework and rationale stated above this study asks two research questions:

R1: Does fidelity of curriculum implementation (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year?

R2: Does the association of one fidelity component with children's gains in science learning outcomes depend on levels of another fidelity component?

Definitions of Terms

Curriculum	A set of guidance in the form of sequenced learning objectives, activities, and materials provided to the teachers to use in the course of a semester or an academic year.
Fidelity of Curriculum Implementation or Enacted Curriculum	The faithfulness with which teachers enact the core components of curriculum in their classrooms, as intended by the authors of the curriculum (Superfine et al., 2015).
Teaching Practice	
Content-Specific Strategies	Planning, use of plans, and adjustments made to meet instructional goals and/or objectives of an academic subject.
Generalized Teaching Strategies	Planning, use of plans, and adjustments made to meet children's developmental needs which can be observed across all types of content areas.
Pedagogy	Theory and method used to teach others.
Components of Fidelity of Curriculum Implementation	
Adherence	Whether teachers delivered the components (e.g., structure, guidance, steps, and script) of the curriculum as prescribed.

Dosage	The prescribed number, duration, or frequency of curriculum lessons teachers used.
Quality of Delivery	The manner in which the teachers used the curriculum using generalized teaching strategies that meet children's developmental needs.

Assumptions and Delimitations

A major assumption of this study is that the curriculum researchers' conception of core components accounts for the critical aspects of teaching practice that will help children learn science. Second, this study assumes that data collected through observation of teachers' instruction in the classroom represents teachers' level of fidelity to curriculum designer and researchers' intended use of curriculum. Another assumption is that the science assessment measured children's science knowledge and skills and that assessment protocols were closely followed by all testers to give equal opportunities without influence to all children who took the assessment.

A delimitation of this study is that although teachers' instructional practice is very complex, this study is based on the theoretical assumption that aspects of practice can be categorized and represented by behaviors that are related to children's learning. This study assumes that in a high-quality academic curriculum, the curriculum designers and researchers communicate the most relevant science content knowledge and pedagogical knowledge with the goal of helping teachers demonstrate the best instructional behaviors. This study will rely on existing conceptions of fidelity and teachers' instructional strategies. Thus, this study will neither attempt to validate nor discredit the specific permutations of fidelity in fidelity instruments and generally accepted categories of teachers' strategies.

Summary

In this chapter, I described the environment in which preschool science curricula would be used by teachers as tools to improve science education for children. I also explained how this study conceptualizes fidelity of curriculum implementation. Although the research on preschool science learning and teachers' fidelity of curriculum implementation is sparse, the resources and expenditures on efforts to support high fidelity implementation of research-based curricula are testaments to the recognition that an excellent curriculum is one that can disseminate high-quality instructional experiences throughout diverse classrooms and schools. When implemented with high fidelity, curricula could help teachers provide children with high-quality learning experiences with purposeful strategies and developmentally appropriate interactions with students. However, as my literature review in the next chapter will show, the evidence of association between fidelity of preschool curriculum implementation and children's learning outcomes is mixed, and even more limited for science curricula.

In the next chapter, I review studies on preschool academic curricula that have examined fidelity and fidelity components. I identify which components of fidelity have been studied in the context of teachers' classroom implementation and outline examples of constructs of fidelity in those studies. The review shows that adherence, dosage, and quality of delivery are the most frequently examined components of fidelity in curriculum research. The studies included in the review had mixed findings on fidelity of curriculum implementation's association with children's learning outcomes. Furthermore, preschool curriculum implementation studies have not yet examined the complex mechanisms of fidelity components in their associations with children's learning outcomes.

CHAPTER 2: LITERATURE REVIEW

Overview

In this literature review, I first describe the limitations of discourse on fidelity of curriculum implementation broadly, then in preschool curriculum implementation studies. After that, I present what existing research on curriculum implementation has shown about the components of fidelity of curriculum implementation. Finally, I propose research questions that are aimed at better understanding the mechanisms of teachers' fidelity in improving children's science learning outcomes.

To understand what is known about fidelity of curriculum implementation and preschool science learning outcomes, I conducted a review of literature published from January 1998 to June 2019 about curriculum implementation fidelity, preschool curriculum implementation and fidelity, teachers' fidelity to curriculum implementation, and teachers' expert practice and children's learning and development. I searched EBSCO with the terms "preschool," "pre-K," "early childhood education," "curricula," "fidelity," and "implementation" to identify relevant articles. From this set of articles, additional publications were identified through a scan of the references in the articles. In addition, I scanned lists of recommended articles automatically generated by the Science Direct database and Mendeley database based on my history of interest in articles. For all articles, I reviewed the abstracts to identify relevant articles, then scanned the reference sections for additional articles that might be included in this review. For the review of classroom context specific literature, I used the following exclusion criteria: interventionists should be teachers with audience being all learners in the classroom, regardless of format (whole group and small group), and curricula should include learning objectives for one or more academic subjects (i.e., language, math, science, and social studies). Although curricula of interest are those designed for an academic year, curricula that were studied for shorter durations were included in the body of literature reviewed. This resulted in my reviewing peer-reviewed journal articles, chapters, reports, and publications from governmental and non-governmental organizations, and conference presentation papers to inform this chapter.

Utility and Functionality of Curriculum in Teaching

Historically, curriculum materials have been used as a part of the reform agenda to improve quality of education; changes in agenda communicated through standards, for example, require changes in teaching practice and materials (Fullan, 2000). Curricula are tools that are designed and developed with the purpose of augmenting aspects of classroom teaching practices in order to improve students' learning experiences and outcomes. Tools are sociocultural artifacts that allow for production and reproduction of socially and culturally based goals across time and space (Vygotsky,1978; Hutchins 1995). Science curricula are artifacts that represent the culturally shared/transmitted beliefs, knowledge, and skills of scientists, science content specialists, pedagogical specialists, education researchers, and curriculum designers. By the using or implementing curriculum, teachers take part in those values, knowledge, and skills of teaching science. Thus, dissemination of curricula facilitates this shared knowledge and practices so that children in those classrooms have the opportunity to experience similar improvements in quality of education.

Curriculum materials are intended to improve the quality of instruction in classrooms; however, their use in the classroom is not always implemented as envisioned by curricular designers. The term *fidelity* is used broadly to mean the degree to which core components that are theorized by curriculum designers and researchers to be most important are implemented as intended by teachers (e.g., Darrow, 2013; O'Donnell, 2008; Century et al., 2010). Research demonstrates that effective early childhood interventions require adequate fidelity of implementation (Domitrovich & Greenberg, 2004). However, even when provided with well-designed curricula, many teachers find it difficult to implement curricular materials in their classrooms with high fidelity across all facets (e.g., Durán et al., 2016). In particular, teaching in the preschool classroom involves more than the guidelines stated in the curriculum. Teachers necessarily need to contextualize and adapt materials for the students in their classrooms (Ball & Cohen, 1996; Penuel, 2014). For example, a preschool teacher who normally conducts her classroom in a manner that promotes free choice and play may need to adapt curricular materials that focus on one learning objective and outlines a specific pedagogical path. Thus, understanding the fidelity of implementation of preschool curricular materials is crucial to understanding the effect of curricular materials on student learning outcomes.

Untangling Fidelity

In preschool curriculum research, teachers' fidelity to curricular intervention remains a nebulous construct that various researchers have represented differently in scope and specificity (Darrow, 2013). Fidelity has been examined as a flexible construct that can be formulated within each study, as a term used to consolidate core components of the implementation as a whole or part, and as few or numerous essential facets of implementation integrity theorized for that study (e.g., PCERC, 2008).

Some of the discourse on fidelity in early childhood interventions has drawn directly from or adapted the conceptualizations of fidelity of school-based health and behavior intervention studies. Recognition of this heritage can help inform the divergent ideas about fidelity and identify the point of departure for building on those constructs in preschool curriculum research. In the broader field of school-based intervention program research, fidelity has been used by some researchers to represent the integrity of prescribed implementation (e.g., Dane & Schneider, 1998; Dusenbury et al., 2003; Greenberg et al., 2005), and by others as an aspect of implementation equivalent to adherence, integrity, and faithful replication and as one of many implementation-related components such as exposure and quality of delivery (e.g., Domitrovich et al., 2010; Durlak, 2010; Nelson et al., 2012).

Similarly, divergent definitions and conceptualizations of fidelity exist in early childhood curriculum research (Darrow, 2013). O'Donnell's (2008) construct of fidelity is often cited in preschool school curriculum research (e.g., Hamre et al., 2010; Mashburn et al., 2016; Phillips et al., 2017; Superfine et al., 2015). O'Donnell (2008) relied on five components of fidelity as derived in two reviews of school-based prevention studies (Dane & Schneider, 1998; Dusenbury et al., 2003): adherence, duration, quality of delivery, participant responsiveness, and program differentiation. Another often cited source of fidelity construct in preschool curriculum research is Odom et al. (2010). Odom and colleagues (2010) utilized the process- and structurally-oriented classifications of fidelity (e.g., Mowbray et al., 2003) in their examination of early childhood curriculum implementation.

Century and colleagues (2010) developed a framework for specifying implementation fidelity that could be applicable to instructional materials programs in multiple subject areas (e.g., teaching guides and student workbooks for a biotechnology center). This project's scope was programs rather than just classrooms, and many types of interventionists rather than just teachers. However, the authors offered a framework for fidelity specific to curriculum programs which is scarce in curriculum studies and which is based their work on written math and science materials.

Building on the five components of fidelity (Dane & Schneider, 1998) and two categories of fidelity (Mowbray et al., 2003), and a content review of written curriculum materials (e.g., teaching guides and student books) of five science and mathematics instructional materials programs, Century and colleagues' (2010) framework proposed four categories: *structural-procedural* which aligned with exposure and dosage; *structural-educative* which is unique to this framework and encompasses the set of knowledge about the subject matter and pedagogies that are meant to be educative for the teachers; *instructional-pedagogical* aligned with quality of delivery; and *instructional-student engagement* aligned with participant responsiveness. They considered adherence and fidelity to be equivalent (Century et al., 2010). This framework has the advantage of addressing categories of communicated curriculum such as a notes section of instructional material that addresses teachers' need for additional content knowledge. Applied to curriculum as defined in the present study, as illustrated above, the majority of the components align with commonly employed fidelity components. Applied to preschool curriculum implementation, this framework presents a possible challenge in separating the resulting enactment based on *structural-educative* components that contain both pedagogy and content in the structural category with *instructional-pedagogical* components that contain pedagogy in the instructional category, as delineation between the pedagogy in the two categories are unclear.

Ruiz-Primo (2005) proposed a matrix for studying fidelity of inquiry-based science curricula that would connect dimensions of curricular programs (i.e., theoretical foundations, curriculum materials, instructional transactions, and intended goals for students) to the five components of fidelity proposed by Dane and Schneider (1998). Within each dimension, critical components for fidelity are outlined. Curricula as defined in the present study fits within Ruiz-Primo's (2005) "curriculum materials*"* that includes "content and activities that have been put together in specific sequence and that can take different forms of documentations" (Ruiz-Primo, 2005, p. 14). Curriculum materials are connected to adherence and exposure. Examples of critical components of curriculum materials are carefully sequenced investigations that are connected to students' prior experience. The *instructional transactions* include teachers' interactions with students that support students' science knowledge such as making connections between lessons, facilitating discussions, and guiding students through scientific processes. These transactions are separated into three categories, of which developing and using scientific knowledge is connected to adherence, exposure, and quality of delivery. The other two categories, *providing learning opportunit*ies and *supporting student learning* are connected to adherence, exposure, quality of enactment, and student responsiveness. An example of critical component of developing and using scientific knowledge is introducing new vocabulary words and reviewing relevant knowledge and skills from prior lessons. Thus, this framework delineated the types of teacher enactment conveyed in the curriculum program in order to determine which aspects of fidelity might be represented in the enactment.

Preschool Curriculum Fidelity

Researchers in preschool settings have measured fidelity in a variety of ways, many without exploration and alignment to other existing conceptions of teacher practice (e.g., Gonzalez et al., 2011; PCERC, 2008). For some researchers, aspects of fidelity, such as adherence and dosage are each alternative means of capturing intended implementation (e.g., Durán et al., 2016). For others, all components of fidelity are important to capture to determine and examine the whole of fidelity (e.g., Darrow, 2013). Still others do not clearly position their conceptualization of fidelity within the constellation of implementation evaluation or historical context of discourse on fidelity (e.g., Durán et al., 2016; Gonzalez et al., 2011).

A large-scale review of preschool curricula embodies this heterogeneous and fluid conceptualization of implementation fidelity. In the PCERC's (2008) review of curricula, participating researchers collected information about the extent to which their materials were utilized in the classrooms. The PCERC's (2008) global measure of fidelity, which asked researchers to give fidelity an overall score—"high," "medium," "low," and "not at all" based on each study's own fidelity measure – reflects a broad or catch-all conceptualization of fidelity wherein fidelity can be defined by the researchers. For example, this conceptualization makes no distinction between two studies that report "high" fidelity when one used adherence and the other study used adherence, dosage, and quality. It also allows for the overall score to be determined using the researcher's judgement regarding what constitutes each score category—for example, whether to give a sum, use a ratio, and what level would be considered medium versus high.

In contrast with this broad conceptualization by PCERC, the researchers of the studies included in PCERC's review used a variety of measures, reflecting a variety in their conceptualization of fidelity. Darrow’s (2013) review of the study-specific fidelity measures used by the researchers who conducted PCERC’s studies showed that conceptions of fidelity components varied. Researchers gathered, with their fidelity measures, information about categories such as the physical environment, activity structures and types, teacher-child interactions, assessment, and family involvement, but without any consistency in the inclusion of categories among researchers (Darrow, 2013). The items in site-specific fidelity instruments ranged from eight to 314 (Darrow, 2013).

Together, these differences indicate a wide range of perspectives on scope and granularity of the measurements that should be required to capture fidelity in preschool curriculum research. Within a broad, inclusive theory of fidelity such as that represented in PCERC’s review, fidelity can be compared and examined across studies because the varying constructs captured in each study are considered equally valid. However, simply adding up or averaging items for an overall score of fidelity may not appropriately represent fidelity in its relationship with children’s outcomes (e.g., Guo et al., 2016; Justice et al., 2015; Piasta, Justice et al., 2015).

Fidelity Examined through its Components

A review of K-12 curriculum research supports the association between fidelity and student outcomes; O’Donnell (2008) found that greater fidelity was significantly and positively associated with student skills or knowledge gains. However, very few published preschool curriculum studies in my review support this association and even fewer preschool curriculum studies have examined the relationship amongst composite mechanisms of fidelity and whether fidelity components predict children’s outcomes; this reveals the need for more research in this area.

Efforts by scholars of preschool instruction concerned with the state of definition and conceptualization of fidelity have focused on identifying and describing the necessary components (e.g., Clements & Samara, 2008; Darrow, 2013). As I will show next, studies of preschool curricula that have examined components of fidelity indicate that a summing of fidelity components may not be representative of the complete picture of teachers’ fidelity in their contribution to children’s learning. I reviewed studies that examined the effects of fidelity and its components on children’s outcomes, and those that considered the relationships amongst the curricular fidelity components of implementation on children’s outcomes.

Very few studies have examined the association of fidelity with children’s outcomes as the result of pre-kindergarten to grade-twelve curriculum implementation and of those that do, very few tried to explain the variability in outcomes in the treated condition (O’Donnell, 2008; Darrow, 2013). Outside of the PCERC studies, even fewer studies with this agenda have been conducted on preschool science curricula, as relatively few studies exist that focus on science education in preschool settings. I identified only two studies on science curriculum that examined fidelity’s association with learning outcomes. In an implementation pilot study with eight teachers who implemented the science and math curricula for one academic year, results suggested that curricular adherence, dosage, and quality of instruction were related to children’s mathematical skills (Kinzie et al., 2015). In another study of social studies and science curriculum, researchers used one overall score for fidelity and did not find that it predicted children's outcomes (Gonzalez et al., 2011). The measures provided by the authors appeared to capture adherence and quality.

In consideration of the relative newness of emphasis on science in preschool at policy and guidance level, I also examined studies on implementation of preschool curriculum on other academic subjects to help determine whether they examined components of fidelity and if so, whether the studies show fidelity components to be predictive of children's learning outcomes and/or related in some way, as to suggest a more complex relationship that their contribution to children's learning outcomes.

Very few studies shed light on which aspects of fidelity are important for children's learning outcomes, however, there is still some indication that quality, adherence, and dosage may be related to children's learning. In curricular interventions that used multiple measures of student outcomes, each of the fidelity components did not consistently predict one or more of the outcomes. For example, Domitrovich et al. (2010) used many outcome measures of socio-emotional, language, and literacy skills. Implementation fidelity—which appeared to measure a combination of adherence and quality under the definitions of this study—significantly predicted an increase in socio-emotional and language skills, while dosage had effect on a subscale of one of the outcome measures (Domitrovich et al., 2010). Similarly, Hamre et al. (2010) found that three of five fidelity component measures were associated with children's learning gains. In the study, preschool students' performance on one or more of four different outcome measures of language and literacy skills were significantly predicted by duration of curriculum activities, quality of delivery, and one of two measures for quality of content specific strategies, but surprisingly, were not predicted by adherence to the curriculum (Hamre et al., 2010).

Three studies examined the relationship among the fidelity components. Piasta, Justice et al. (2015) examined the largest number of fidelity variables found in my review. In their study of a 30-week implementation of a supplemental language and literacy curriculum, *Read it Again!* (*RIA*), they examined preschool teachers' adherence, exposure, quality of delivery, and participant responsiveness at up to three time points. They found dosage had a significant association with children's scores on one of four assessment measures of language and literacy, and language-specific strategies had a significant association with two of the outcomes. Of the four types of language and literacy learning gains examined, print knowledge gains were significantly and positively associated with dosage, i.e., the time teachers spent on lessons and employing scaffolding. Quality measured through teachers' scaffolding was also positively and significantly associated with phonological awareness. Adherence and participant responsiveness were not associated with learning gains. Piasta, Justice et al. (2015) found that some of the fidelity component measurements showed a significant correlation (adherence and quality, and adherence and participant responsiveness) but no discernable patterns emerged among the different fidelity components. This study represents a notable step for early childhood academic curriculum, in exploring fidelity's multiple conceptual components and their association with children's learning, even though the research was focused on language and literacy and curricular enhancements. Guo et al. (2016) examined adherence and dosage (as one fidelity component), participant responsiveness, and program differentiation in their relation to children's learning outcomes. Guo and colleagues (2016) found that only program differentiation significantly mediated the treatment effect on children's outcomes suggesting that it explains the treatment effect. They also found that a three-factor model showed significantly better fit than the one-factor model, giving support to the conceptualization of separate fidelity components rather than using an overall score.

Justice et al. (2008) examined the relationship between adherence to procedures provided in the *MyTeachingPartner Language and Literacy* curriculum and one quality dimension. The researchers found that fidelity to the preparation and structure of activities (e.g., including all the materials listed in the activity plan and preventing distractions) was not significantly related to either of the two quality measures used in the study. However, fidelity to how the lesson should be delivered (e.g., using the script provided in the activity plan and encouraging participation) significantly predicted one of the quality measures.

In contrast to the diversity in fidelity components examined by Piasta, Justice et al. (2015), and the multi-dimensionality of fidelity supported by Guo et al. (2016), a study of *Pre-K Mathematics Curriculum*, Klein, Starkey, Clements, Samara and Iyer (2008) used one fidelity measure score reflecting a range of fidelity components of adherence and quality of delivery. This study did not examine fidelity in its relation to children's outcomes. Klein and colleagues (2008) reported mid- to high fidelity levels by teachers in the treatment group and found that children in the treatment group had significantly higher gains in mathematics scores than children in the business-as-usual control group. Since fidelity was not examined in relation to children's outcomes, it is difficult to determine whether and how fidelity to the curriculum might have supported the higher outcomes in the treatment group.

Finally, numerous studies reported moderate to high implementation fidelity but no significant associations between fidelity components and children's outcomes. In another study of preschool classrooms, the authors of the *Building Blocks* curriculum (Clements & Samara,

2011) and an integrated *Building Blocks* and *Preschool Mathematics Curriculum* (Klein, Starkey, & Ramirez, 2002; Sarama et al., 2006) found that the total fidelity score of 61 items—which included adherence, dosage, and quality—correlated with children's gains in mathematical knowledge but that the relationship was not significant. Both trials' teachers had moderate to high levels of fidelity.

In summary, individual components of fidelity have shown mixed indications in their relationship to one another and in their effect on learning outcomes. More studies should examine curriculum implementation with a multi-component model of fidelity wherein each component may have different values and have different effects on learning outcomes. Only one study on curriculum examined the correlation amongst fidelity components (Piasta, Justice et al., 20015) and one study examined fidelity components' dimensionality in relation to children's outcomes (Guo et al., 2016). These findings have to be understood with the caveat that interpretations about fidelity beyond individual studies has been made difficult by the flexibility and ambiguity in what should be considered core components of preschool curriculum enactment.

So far, I have illustrated how in discussions of fidelity in curriculum research, limited headway has been made toward tailoring the components and descriptions of fidelity to fit the context of preschool curriculum research. I have shown how operationalized fidelity is contextual but exploration of the concept of fidelity has been largely built on borrowed concepts and definitions of fidelity from school-based prevention intervention research, without explicit contextualization of a fidelity framework to preschool teachers' expertise and practice. The applicability and value of definitions and constructs of fidelity from health and prevention programs to the preschool curricular intervention context have been observed and discussed in curricular intervention research (e.g., Hamre et al., 2010; O'Donnell, 2008; Odom et al., 2010). Although conceptualizations from school-based interventions can be applicable and useful, school-based intervention literature tends to focus on program theory in the organizational environment, while curriculum studies should focus on teacher practice (O'Donnell, 2008). This critical distinction can be used to take the components derived from prevention literature and develop our understanding in the context of teaching in preschool classrooms.

I have shown in my review that whether examined with a broad lens, or through its components, it is uncertain whether fidelity and its components are related to children's learning outcomes. No study has examined whether fidelity components have a relationship to one another that strengthens or weakens their relationship to children's outcome. Studying fidelity components and their effects on outcomes could help to explain which components are critical in order to produce desired effect on students' learning. However, not enough studies in curricular interventions have examined fidelity with the goal of informing which components are critical and in which instructional contexts. An important next step for preschool curricular intervention studies would be to more closely examine components of fidelity, beyond reporting it, to inform if and how the curriculum's theorized core components influenced children's learning outcomes when implemented by teachers in their classrooms.

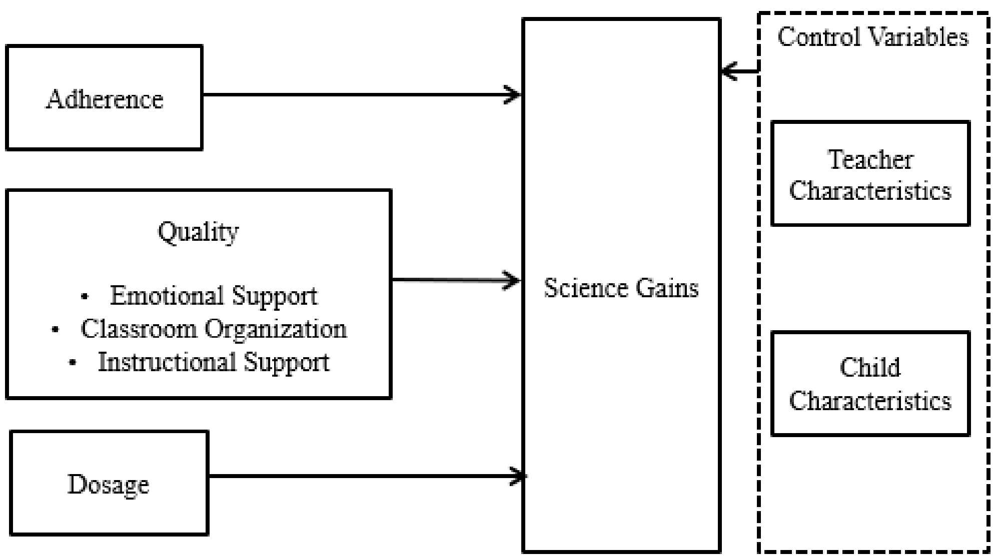
Research Questions and Hypotheses

This study draws on the data collected as part of the *MyTeachingPartner-Math/Science* (*MTP-M/S*) curriculum implementation study. I used the data collected from the teachers, children, and video-based classroom observations from the classrooms where the teachers used the *MTP-Science* curriculum. More details about the *MTP-M/S* study and methodology are provided in Chapter 3. My dissertation focuses on the following research questions, with my associated hypotheses:

R1: Does fidelity of curriculum implementation (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year? Based on research on implementation suggesting that fidelity of implementation of early childhood curricula is related to learning outcomes (Domitrovitch et al., 2010; Hamre et al., 2010; Kinzie et al., 2015, Piasta, Justice et al., 2015), I hypothesize that the individual fidelity components will be positively associated with children's science learning. The theoretical model of the first research question is represented in Figure 6.

Figure 6

Research Question 1, Hypothesis 1: Additive Effects of Adherence, Quality, and Dosage



R2: Does the association of one fidelity component with children's gains in science outcomes depend on levels of another fidelity component? I proposed three hypotheses for two-way interactions. First, I hypothesized that the relationship between teachers' adherence to lesson design on children's gains in science learning outcome is moderated by quality of delivery. More specifically, when teachers enact lessons with higher quality teacher-child interactions, the positive effect of their adherence to lesson design on children's science gains is stronger than when they teach with lower quality. Similarly, I hypothesized that the relationship between the dosage of science lessons that teachers implemented and children's science gains will also be moderated by quality of delivery, such that the positive effect of the quantity of science lessons teachers implement on children's science gains is strengthened when teachers deliver lessons with higher quality of delivery and weakened when teacher deliver lessons with lower quality of delivery. Finally, I hypothesized that the relation between teachers' adherence and children's gains in science learning outcome is moderated by dosage of science lessons, such that the positive effect of teachers' adhering to the intended lesson design on children's science gains is stronger when teachers implement more lessons and weaker when they implement fewer lessons. The theoretical relationships for the second research question are in Figures 7, 8, and 9.

Figure 7

Research Question 2, Hypothesis 1: Quality Moderates Adherence

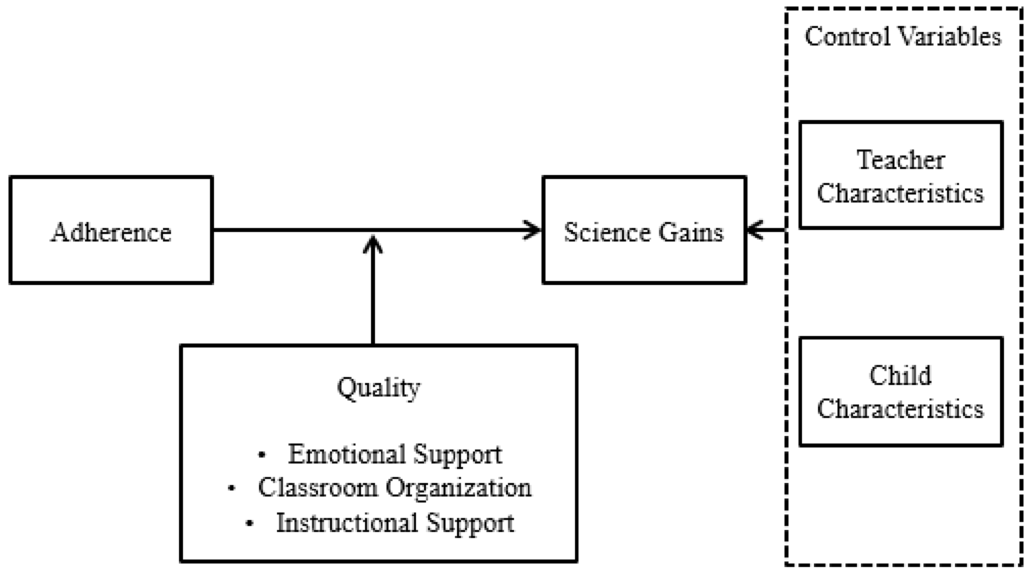


Figure 8

Research Question 2, Hypothesis 2: Quality Moderates Dosage

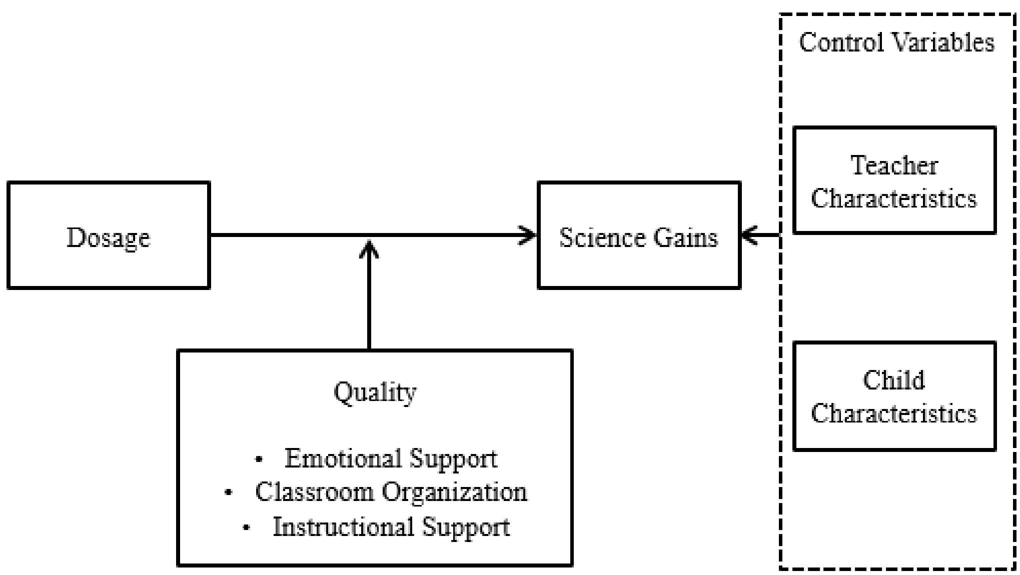
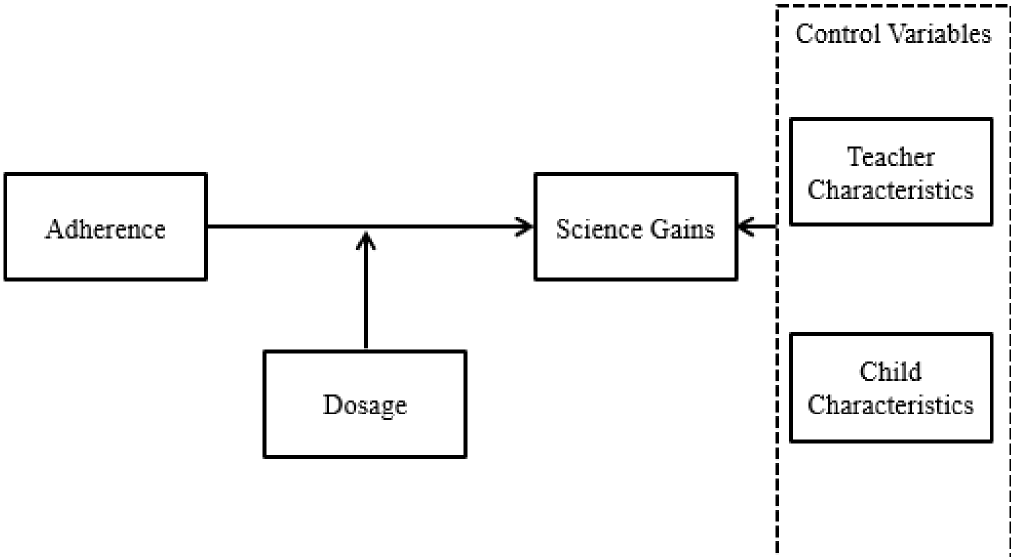


Figure 9

Research Question 2, Hypothesis 3: Dosage Moderates Adherence



As described in the next chapter, quantitative analyses of multiple linear regression models were conducted to test the above questions and gain a deeper understanding of the components of fidelity and the associations amongst them in their relationship to children's learning outcomes.

CHAPTER 3: METHODS

Study Description

This study examined the extent to which implementation fidelity relates to preschool children's science learning, using data from the *MTP-M/S* randomized controlled curricular intervention. Given the dearth of published literature that has examined how components of fidelity of implementation are related to children's science outcomes and how those fidelity components might interact with each other to effect children's science outcomes, my research questions sought to examine those relationships. Specifically, I explored the following research questions:

R1: Does fidelity of curriculum implementation (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year?

R2: Does the association of one fidelity component with children's gains in science outcomes depend on levels of another fidelity component?

To answer these research questions, I used multiple linear regression models. To explore the first research question focused on the associations between adherence, quality of delivery, and dosage and preschool children's science outcomes, multiple regression models with all three fidelity components in each model were used to test the main effects of each predictor. The second research question was addressed by examining the interactive effects of 1) adherence and quality of delivery, 2) dosage and quality of delivery, and 3) adherence and dosage. For this second research question, two-way interaction models were used in multiple regression analysis.

The dependent variable in all models was children's scores on a science assessment, which assessed life science and earth and physical science knowledge and skills. Independent variables of interest were the teachers' adherence to the curricular design, the dosage of curriculum implemented by teachers, and the quality of delivery with which the implementation occurred. A more detailed plan for hypotheses testing and associated models are described in the Data Analysis section.

I examined the extent to which components of teachers' fidelity of curriculum implementation matters for children's science outcomes. I focused on how components of implementation fidelity may interact, so that one component's association with children's science learning outcomes may be moderated by another component, in the context of an efficacy trial of a pre-k science curriculum implementation.

I begin by presenting a description of the *MTP-M/S* intervention study from which I drew the data for my study. Then, I present a description of this study's participants, data collection procedures, and measures. After that, I present how I carried out the data analysis, including data preparation and examination, description of covariates, descriptive analyses of key variables of interest, missing data imputation, data screening for multiple regression analysis, and data analyses procedure. Finally, I explore possible limitations of the study, including power, and bias from validity, and quality.

Data Source: *MTP-M/S* Intervention Study

As my study relied on a subset of data from the *MTP-M/S* intervention study, I provide a description of the nature and design of the larger *MTP-M/S* intervention study implemented during the 2009-2010 academic year. My dissertation used the data collected from the teachers and students in classrooms in which teachers implemented the *MTP-M/S* curricular activities.

MTP-M/S Intervention Study Design

With a grant from the U.S. Department of Education, researchers at the Center for Advanced Study of Teaching and Learning (CASTL) developed new math and science preschool curricula and conducted a small randomized field trial on the feasibility and effects of implementing the new curricula in a single school district (Kinzie et al., 2014; Whittaker et al., 2016).

In the *MTP-M/S* study, random assignment was implemented at the school level, in order to avoid the contamination that could occur when teachers assigned to different conditions work together and share information at the same school. Stratified proportionate randomization by number of participating teachers in the school was conducted in order to assure that the three experimental groups (one control group and two treatment groups) had balanced number of teachers to the extent possible. This ensured that all of the schools had equal chance at being selected for the control and treatment groups.

Teachers who implemented *MTP-M/S* curricula were assigned to *Basic* or *Plus* groups and received different levels of professional supports based on the group designation. Each teacher in the *Basic* and *Plus* groups received the same lesson plans, monthly activity checklists, send-home newsletters, and activity materials. As I describe below in *MTP-Science* Curriculum, the lesson plans contained substantial teacher supports. Teachers were asked to use the lesson plans and materials when implementing lessons. The research team provided the majority of the activity materials, which included manipulatives, books, and various other supplies typically not found in abundance in classrooms (e.g., soil).

Teachers in all groups attended a one-day workshop during which they were oriented to the study. In addition to the physical materials and the orientation workshop, teachers in the *Plus* group received professional development support. This support is described under *MTP-S* Professional Supports.

Next, I summarize previous findings from the original *MTP-M/S* intervention study, then go on to describe the *MTP-Science* curriculum materials and teacher supports.

MTP-Science (MTP-S) Curriculum

The *MTP-S* curricular activities were informed by national and state standards for kindergarten and state standards for pre-k. The curricular design team developed trajectories of science learning for the pre-k year, based on K-2 Benchmarks from the American Association for the Advancement of Science (AAAS, 1993) and K-4 National Science Education Standards (NRC, 2006). Broadly, the *MTP-S* curricular activities covered topics in life science and earth and physical science. A list of weekly activities and titles is provided in Appendix A. Prior to the implementation in the 2009-2010 academic year, the *MTP-S* curricular design and implementation was informed by an iterative design process first involving five pre-K teachers in two local schools, followed by a year-long pilot implementation with eight teachers who used the curriculum in their classroom (Kinzie et al., 2015).

The *MTP-S* curriculum included two science activities per week for the academic year. Each activity was designed to take approximately 20 minutes for a teacher to implement with children. Each activity had a lesson plan that the teachers were to follow. Each lesson plan contained the following sections: 1) *Get Ready*, 2) *Engage*, 3) *Investigate*, 4) *Discuss*, 5) *Extend*, and 6) *Make it Work*. A sample activity guide sheet is provided in Appendix B.

The *Get Ready* section contained structural information for carrying out the activity including learning objective(s), important vocabulary words, activity format (small group or whole group), required materials, and necessary preparation. The next section, *Engage*, contained steps to draw children's attention and ready them for the activity. This section included first having children recite a short chant to signal that they will employ their senses for studying science; then what teachers should do, say, and ask to get children to think about the topic of the activity; and lastly, a statement or question that captures what the children should think about or do next. The main body of the activity, *Investigate*, provided a sequence of component activities such as book reading and experimenting, accompanied with explanations that should accompany them and questions that teachers should pose to promote active participation and thinking. Next, the *Discuss* section provided ways for the teacher to close the lesson by summarizing and/or reviewing what children learned during the activity. Finally, the last two sections, *Extend*, and *Make it Work* provided recommendations for extensions into other activities across the classroom day, and scaffolding methods teachers can use to support children who need simplification or more challenge to engage in the activity.

MTP-S Professional Supports

In addition to the curricular materials and one-day orientation workshop all treatment group teachers received, the *Plus* group received access to the *MTP-M/S* website and attended seven face-to-face, two-and-a-half hour workshops about science, mathematics and quality of delivery. Of the seven workshops, all included quality of delivery related topics (*instructional learning formats*, *concept development*, *quality of feedback*, and *language modeling*) and four included science topics (worms, floating and sinking, magnets, and plant life cycle). The *MTP-M/S* website provided over 130 videos, of about two or three minutes in length, that contained model implementations of the mathematics and science curricular activities. The implementation videos and 150 additional videos on quality teacher-child interactions helped focus teachers' attention on best pedagogical practices and how children construct knowledge (especially alternative ideas), and aimed to help teachers develop understanding of mathematics and science concepts.

In the *Plus* group workshops, teachers worked on developing science concept knowledge, an inquiry model of learning, and how to improve their quality of interactions with children based on dimensions of the *Classroom Assessment Scoring System (CLASS*, Pianta, La Paro, & Hamre, 2008; described in Quality section below). Workshop participation also included familiarization with the website and how to integrate the practice teachers observed in the videos into classroom practice. In these workshops, teachers engaged in self- and peer review as well as error analysis of pedagogy and group discussions, identifying how they would improve their classroom practices.

Additional Supports

In addition to the lesson plans and materials necessary to implement lessons, monthly knowledge and skills checklist was provided to treatment group teachers to help them track the learning of the individual children in their classrooms. Treatment teachers were also provided monthly family newsletters to send home to children's parents or caregivers. The newsletter contained descriptions of what children were learning and suggestions for activities they could engage in at home to strengthen children's learning. The use of these additional supports was encouraged, but not studied.

Lastly, teachers were provided with a monthly Activity Checklist with which they could keep track of completed *MTP-M/S* curriculum activities. This checklist was intended to encourage compliance and record teachers' use of the curricula.

Previous Findings

Some notable findings from the *MTP-M/S* field trial have been published. Researchers found significant effects of the curriculum on the quality of teachers' interactions with children—in their quality of delivery and in using mathematics and science teaching strategies (Whittaker et al., 2016). Although the implementation of curriculum did have a significant effect on children's math outcomes, it did not have a significant effect on children's science outcomes (Kinzie et al., 2014).

In published *MTP-M/S* papers, authors reported that there was no overall effect of treatment (vs. control) on children's science scores. They reported that no significant difference was found for children's gains in earth and physical science (*EPS*) scores between the teachers who used the curriculum and were given access to additional web-based professional support, and the teachers who used the curriculum and were not given access to the web-based support. However, for children's life science (*LiS*) gains, the researchers observed a significant difference between the teachers who received different levels of professional support (Kinzie et al., 2014). Thus, the level of professional development support teachers received may be associated with children's science learning outcomes. My research sought to build on these findings and examined the main effects and interaction effects of components of teachers' curriculum implementation fidelity on children's science learning outcomes. Next, I describe the *MTP-Science* curriculum and professional supports provided to the teachers who used the curriculum.

Participants

In the original *MTP-M/S* study, 42 teachers were recruited from a single district. Eleven teachers in the business-as-usual group implemented their existing curriculum; 17 teachers in the *Basi*c group received the curricula materials; and 14 teachers in the *Plus* group received the curricula materials, plus additional professional development support described above. The participants in my study included the teachers who implemented the research curricula (i.e., *Basic* and *Plus* teachers) and children in their classrooms. Following the description of the participants in my study, the Procedures section includes details on how teachers were recruited and children were selected for participation by the *MTP-M/S* researchers.

Teachers

In total, 31 teachers (N = 31) who implemented the *MTP-S* curriculum were included in my study. All teachers taught in schools with Title I designation, which means greater than 40% of students in the schools were from low-income families. Teachers ranged in age from 27 to 65 years (M = 45.11, SD = 10.53) with two to 32 years of experience working with pre-kindergarten children (M = 8.05, SD = 7.02). All teachers held a bachelor's degree, which was a minimum requirement to teach in state-funded programs. About half of the teachers (52%) held a master's degree. Their classroom sizes ranged from 12 to 18 students (M =16.21, SD = 1.55). Table 1 presents descriptive information for teachers and their classrooms.

Table 1

Teachers and Classroom (n = 31) Characteristics

	N	Missing	*M*	*SD*
Teacher characteristics				
Teacher's age	27	4	45.11	10.53
Holds a Master's [a]	31		0.52	0.51
Experience with preschoolers [a]	28	3	8.05	7.02
Classroom characteristics				
Plus [a]	14			
Basic [a]	17			
Number of students in class	28	3	16.21	1.55

[a] Covariate.

Children

From the classrooms of teachers who implemented the *MTP-S* curricula, 328 children were selected for direct assessment and were included in this study. The number of selected students in classrooms ranged from eight to thirteen (M = 10.58, SD = 1.06).

As the students were attending public schools, they were all assumed to be eligible for kindergarten the following year. The average age of children in the treatment classrooms was 4.6 years (range: 2.92-5.71, SD = 0.322). A little over half of the children were female (53%). The selected children were predominantly Black/African-American (67%). A smaller percentage of children were White/Caucasian (24%), with less than 10% of children being of other ethnicities. A range of family demographics were represented in selected children. Maternal education levels ranged from completion of less than eighth grade to holding a graduate degree. The income-to-needs ratio (M = 1.3, SD = .92) ranged from extreme poverty (6% of necessary income) to well-above the poverty line (4 times the necessary income). Thirty-nine percent of children were from families that lived in poverty, with a family income-needs ratio of less than one as defined by the

United States Census Bureau (2004). Table 2 includes descriptive information about the children in this study.

Table 2

Child (n = 328) Characteristics and Covariates

	N	Missing	*M*	*SD*
Child characteristics				
Student age	301	27	4.61	0.32
Family income/needs ratio [a]	292	36	1.30	0.92
Maternal education [a,c]	314	14	4.25	1.57
Gender of student	321	7	0.53	0.50
Male	150			
Female	171			
Ethnicity	322	6		
African-American	220			
Caucasian	77			
Other	25			

[a] Covariate
[b] Dependent Variable
[c] 1 = 8th grade or less, 2 = some high school but no diploma , 3 = high school diploma or equivalent, 4 = high school diploma or equivalent and technical certificate, 5 = some college but no degree, 6 = associate's degree or two-year degree, 7 = bachelor's degree, 8 = graduate degree.

Procedures

Below, I provide the participant recruitment, selection, and data collection procedures for the *MTP-M/S* intervention study. The researchers collected data through demographic surveys, video tapes of science activities submitted by teachers, and direct assessments of children.

Study Participant Recruitment Procedure

Teachers were from a state-funded pre-kindergarten program within a single district in a mid-Atlantic state. All pre-kindergarten teachers in the district were sent a flyer with a description of the project and a form that they could return indicating interest of "yes" or "maybe." A research team member contacted teachers who responded positively to the request to answer questions about the research project and participation. Teachers who agreed to participate signed a consent form and filled out a survey that included demographic information about themselves and their classroom.

From the classrooms of teachers who agreed to participate, children were selected to participate in the study from a larger pool of children whose parents gave consent. First, in order to obtain a parent or guardian's permission for children to participate in the study, the research team drafted and printed copies of consent forms for the teachers to distribute. The consent form contained an explanation of the research project and the scope of children's participation. Teachers were asked to send the consent form home in students' backpacks and follow up with those who did not respond. Of 578 parents and guardians of all possible child participants, 529 (94%) gave consent for their children to participate (Kinzie et al., 2014). Parents and caregivers who consented were also asked to fill out a form with demographic information about the children and families, sent home by teachers. Children who had an Individualized Education Plan (IEP) other than for speech, or had limited English proficiency, were not eligible to participate.

From this pool of eligible children, the research team randomly selected ten children per classroom to participate in direct assessments. If a selected child was not available due to absence or disenrollment, then another child was randomly selected from the pool of eligible consented children from the same classroom.

Direct Science Assessment Procedure

Children's science knowledge and skills were assessed in the fall and spring. Trained data collectors, who were blind to study group assignment, conducted the assessments individually with selected children in a quiet, private location away from the classroom. The data collectors had completed two full days of training on assessment administration and scoring, building rapport with children, and encouraging their persistence. As part of the training, the data collectors administered four practice assessments and the data they collected were examined by the research team for completeness and accuracy.

The science assessments were administered with a battery of other assessments. Each selected child engaged in a battery of assessments for two 30-minute sessions with at least 30 minutes of time in between the sessions as a break, to avoid fatigue for children. If during the assessment, a child indicated that he/she did not know the answer, the examiner encouraged guessing. If after encouragement, the child still did not give an answer, the test administrator marked the response as incorrect and moved on to the next question. Children were not allowed to return to a skipped question or to change their answers to an already answered question after moving on to the next question.

Students selected for assessment were given a book for their participation after the assessment session. After all of the selected students completed the direct assessments, the remaining children in each classroom also received a book. If in the fall, a selected child was absent on the test day, an alternate child was randomly identified to replace the originally selected child. In the spring, if a selected child was absent, the data collector returned on a different day to administer the test to that child.

Videotape Observation Procedure

Tape Submission

All teachers participating in the study received blank videotapes, a video recorder, and a tripod that could be set up in his/her classroom. The research team asked the teachers to video record and submit all science and math activities that they implemented, using one tape for each activity. Teachers were provided postage-paid envelops and asked to submit a target of 15 math and science videotapes per month or 135 across the year, supplementing with videotapes of other curricular activities if they did not implement 15 *MTP-M/S* activities; teachers submitted an average of 85 math and science tapes across the year (*M* = 84.51, *SD* = 42.83) (Whittaker et al., 2016).

Science Tape Selection

In order to capture data that are adequate representations of teachers' practice and also limit cost, time, and resources required to code videos, one science tape from each teacher was randomly selected from each teacher's submitted tapes from the months of September, October, November, February, March, and April for observational coding. Those six months were chosen because of the low number of tapes submitted in December, January, and May—months during which there are long school breaks and additional responsibilities for teachers at the end of the academic year. The selection of the tapes alternated between small and whole group activities for each of the six months (i.e., if for a teacher, a small-group science activity was selected for September, a whole-group science activity was selected for October). On average, five science activities were coded per teacher (M = 4.9, SD = 1.70, range: 1-6). The same set of tapes with science activities selected (N = 93) were used to double code for adherence and double code for quality of delivery using the *Classroom Assessment Scoring System* (Pianta, LaParo, & Hamre, 2008). I calculated the coder reliability statistics for observation measures for science activity tapes.

Adherence Coding

For curricular adherence coding, a master coder and three coders were trained by coding a series of tapes from a previous study until the three coders matched the master coder's scores. The master coder was a member of the research team who was deemed to be more familiar with the typical preschool classroom conditions and the adherence measure and had experience with coding from video observations. Coders used an internally developed manual for reference during observation, recording observation notes on paper forms before scoring.

Coders achieved a high degree of interrater reliability on the total *MTP-S* adherence score (ICC = .95), which indicates a high level of agreement amongst the three coders. In addition, interrater reliability analysis at the item level was performed using the Kappa statistic to determine consistency among raters for the five dichotomous items and interclass correlation coefficient (ICC) for the nine ordinal items. The interrater reliability for items ranged from moderate to excellent for all items (range: Kappa = .72-.83, ICC = .91- .97), indicating good to excellent agreement among the coders at the item level.

CLASS Coding

Six coders (observers) were trained to evaluate and code the selected video recordings using the *CLASS* (Pianta, La Paro, & Hamre, 2008), measure of quality of delivery and the adherence instrument (described next). All the *CLASS* coders were certified as reliable by Teachstone. Teachstone is the organization which owns the copyright to the *CLASS instrument* and provides *CLASS* training and certifications for observers and trainers.

To receive the observer reliability certification, coders must attend a two-day *CLASS* Observation Training during which they receive the coding manual, watch master coded video clips of preschool classrooms, and calibrate their scores with master codes. After the training, each coder takes an online *CLASS* Reliability Test during which they must watch and score five video clips. To pass the certification test, the coder must score within one point for each dimension for 80% of the ten dimensions coded per observation, and not score unreliably in any one dimension more than two out of five observations. Each coder has up to three opportunities to pass the test. If the coder does not receive a passing score, then the coder must attend another two-day training session before taking the test again.

Coders used the *CLASS* coding manual for reference during video observations, recording observation notes on paper forms before scoring on each measure. Based on observer reliability certification rules, acceptable reliability standard for *CLASS* is coder agreement within one point in either direction (i.e., a score of one or three on a given dimension is in acceptable agreement with a master-coded score of two) (Teachstone, Observer Certification FAQ). The authors of the measure, Pianta, La Paro, and Hamre (2008) also recommend using percent-within-one analysis, with .8 (i.e., 80%) being the acceptable average for inter-rater reliability.

Although some studies use percentage of agreement for coder reliability for the *CLASS* measure, percentage of agreement is an inadequate measure of interrater reliability because it does not account for expected disagreements, range of scores available, and the population of data scored (Krippendorff, 2012). The intraclass coefficient (ICC) for the domain level average score (*Emotional Support* ICC = .73, *Classroom Organization* ICC = .49, *Instructional Support*

ICC = .63) were acceptable—based on widely used cutoffs provided by Cicchetti (1994)—indicating components of quality of delivery were rated similarly by coders.

Measures

My study drew upon the following measures employed in the *MTP-M/S* curricula study. I describe the measures, score calculations, and data screening and examination undertaken prior to multivariate analysis conducted to test the hypothesis. Internal consistency values were calculated for all of the measures.

Composite Science Score

Composite scores that represent science knowledge and skills at two time points, fall and spring, were calculated from two science measures from the *MTP-M/S* curricula study. The two science measures, the *Life Science Assessment* (*LiS*) and *Earth and Physical Science Assessment* (*EPS*) were created by the *MTP-M/S* research team to align with national and state pre-kindergarten science learning standards. They were designed to test children's factual and conceptual understanding of the biological world and the physical world (Kinzie et al., 2014). These measures employed questions and response options illustrated by images and photos and/or materials that students manipulated in response to forced choice items and card sorting items.

The *Life Science* measure (*LiS*) was composed of 51 items with one point each on the topic of living and non-living things, plants and animals, body parts and functions, plant biology, animal behavior during day and night, and animal families. Primary types of *LiS* items involved forced-choice response selection, and card sorting. *Earth and Physical Science (EPS*) topics included scientific tools, weather, temperature, material composition, motion, and floating and sinking. *EPS* was composed of 50 questions, worth one point each. As with the *LiS* measure, the primary types of *EPS* items involved forced-choice response selection and card sorting. The purpose of the fall assessment was to establish baseline scores for children's knowledge and skills against which change in children's science learning in the spring could be examined. A list of science assessment items is provided in Appendix D. The *LiS* and *EPS* assessments had moderate and significant levels of correlation in the fall and in the spring. The fall *LiS* and *EPS* scores were significantly correlated, ($r = 0.51$, $p < .01$). The spring *LiS* and *EPS* scores were also significantly correlated, ($r = 0.58$, $p < .01$). The moderate significant correlation between the Life Science scores and Earth and Physical Science scores indicated that the two scores may measure some similar elements of science knowledge. I determined that this moderate correlation alone (versus sometimes the more desirable high correlation) did not threaten the validity and reliability of the composite to warrant using two separate science scores. Other considerations are explained below.

Since this study set out to examine children's science knowledge as a whole given teachers' implementation of one science curriculum, one composite score to represent children's science knowledge and skills was calculated from the two science scores (*LiS* and *EPS*). I created one composite science score by first taking the mean of all items for each measure, then standardizing each of the two scores, and finally summing the two standardized scores. The main rationale for using a composite score instead of two separate science scores as dependent variable in this study was to simplify the interpretation of findings, given there is no basis to support the idea that teachers' fidelity of curriculum implementation would have different effects on each of the two science scores. In addition, separating the observation data into *LiS* and *EPS* lesson would have drastically lowered the number of sample tapes coded for each science domain, thus limiting the adherence and quality of delivery measures' ability to represent teachers' levels of fidelity of a particular domain of science instruction over an academic year.

Internal consistency reliability estimates were computed with Cronbach's alpha from direct assessment scores. *LiS* had good internal consistency in the fall ($\alpha = .76$) and spring ($\alpha = .77$). *EPS* also had good internal consistency in the fall and ($\alpha = .77$) in the spring ($\alpha = .82$). The combined measure with 101 items had good internal consistency in the fall ($\alpha = .85$) and in the spring ($\alpha =.88$).

Fidelity Measures

Adherence

The adherence measure was developed by the *MTP-M/S* researchers to reflect the degree to which teachers' implementation aligned with the core components of the curricula that are theorized to influence children's ability to learn from the lessons. It consisted of 14 items measuring activity completion, use of materials, engagement of children, content coverage, supporting cognition and language, and instruction of objective and content. The first five items were scored on a scale of 0 to 1, 1 point for behavior being present and 0 for not being present. For example, activity was a dichotomous item from the adherence measure and was conducted in a specified format (whole group, small group; Item 1).

The next nine items were scored on a scale of 1 (never) to 4 (all of the time). Two examples of items with four-point rating scale were:

- Children use the specified math and/or scientific language in their comments (from "Use the Lingo"). (Item 7)

- Teacher elicits children's observations and explanations for the activity, during available opportunities. (Item 8)

The complete measure and accompanying administrative explanations are provided in Appendix C.

The total adherence score from the sample of selected science activity tapes for each teacher was calculated by standardizing the average item-level score across the tapes and taking the mean of the 14 items to accommodate the mix of dichotomous and ordinal items. The internal consistency of adherence measure for science activities was determined to be high (α = .90). Item nine had no variance, with all teachers receiving the highest score on the item, and therefore, was not included in the item correlation calculation. The items were all positively correlated. Item 14 had a weak relationship with the composite (r = .24) and all other items had moderate to high correlation (r > .31). After reviewing item 14's content validity, the decision was made to keep the item.

Dosage

Another aspect of fidelity is the dosage—the number of times that teachers implemented the science curriculum. Since each tape contained one activity, the number of science tapes submitted by teachers represented the number of times the teacher exposed children to science curricular activities.

As previously mentioned, *MTP-M/S* researchers also collected complementary data from the monthly activity checklist on which teachers could indicate the activities they implemented. Of the 66 possible *MTP-S* activities that teachers could have taped and submitted, they submitted a range of two to 62 science tapes (M = 40.13, SD = 21.07).

I also considered using the Monthly Checklist as a dosage measure. Checklists represented 1638 science activities, which was 79% of the possible 2064 science activities that could be implemented by 31 teachers. The 1638 records represented data recorded when teachers submitted Monthly Checklists. This dataset was examined in order to determine whether it would be an appropriate source of measure for dosage. The following figures indicated that the checklist did not fully serve its intended purpose:

1. 67% of activities were indicated as implemented and teachers submitted corresponding tapes;
2. 14% of activities were indicated as not implemented and teachers did not submit corresponding tapes;
3. 4% of the activities were marked as implemented but no tape was submitted;
4. 1% of the activities were marked as not implemented but teacher submitted tapes; and
5. 14% of the activities were missing implementation information on the Checklist.

Checklist data cannot be used to account for activities that were implemented when they were not submitted, in the same way that the count of submitted tapes does not provide information about instances when activities were implemented but no tape was submitted. The Monthly Checklist could have provided additional dosage information; however, the percentage of missing and incorrect data suggested that the Checklist did not provide reliable information that could complement the information provided by count of science tapes.

Thus, only the count of science tapes submitted was used as a dosage measure. However, the count of video tapes submitted as a measure of dosage also has possibilities for errors. One example of a possible error is that teachers' self-perception of their level of adherence and quality of delivery may have prompted teachers with higher self-perception of fidelity to videotape and submit more tapes and conversely, teachers with lower self-perception may have submitted fewer tapes (Kinzie et al., 2015).

Quality of Delivery

The quality of teacher-child interactions during the science activities was measured with the *CLASS* instrument (Pianta, La Paro, & Hamre, 2008). The *CLASS* measures teacher-child interaction quality with ten dimensions, which have been conceptualized to represent three major domains of teacher-child interactions: *Emotional Support* (*ES*), *Classroom Organization* (*CO*), and *Instructional Support* (*IS*) (Pianta, La Paro, & Hamre, 2008). Large-scale studies have shown that *CLASS* scores are associated with academic and social outcomes (Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008; Burchinal et al., 2010).

Each dimension is scored for specific behavioral markers of teacher behavior, and also reflects the children's behavior in the classroom. Each dimension has a seven-point scale, with a score of one indicating lowest levels of quality for that dimension, except for *Negative Climate*, which is given a 1 when there is absence of behavior that contributes to negative climate (*Negative Climate* is reverse-coded after scoring). *Emotional Support* (*ES*) is composed of four dimensions. *Classroom Organization* (*CO*) and *Instructional Support* (*IS*) are each composed of three dimensions.

For each selected video, the domain level scores were calculated by averaging the component dimension scores. For *ES*, positive climate, reverse-coded negative climate, teacher sensitivity, and regard for student perspectives scores were averaged. For *CO*, behavior management, productivity, and instructional learning format scores were averaged. For *IS*, concept development, quality of feedback, and language modeling were averaged. Then, domain level scores from all selected tapes were averaged at the domain level, producing an aggregate score representative of each teacher's practice in regard to *ES*, *CO*, and *IS.* Internal consistency was good for *IS* ($\alpha = .78$) but was below typically acceptable levels for *ES* ($\alpha = .67$) and *CO* ($\alpha = .63$). Low levels of internal consistency suggest that the items within each domain do not correlate well and perhaps do not represent the same latent variable. However, internal consistency is not the only means of gauging whether a domain is valid.

Existing studies have used the three domains of interaction in examination of their relationship with children's developmental and learning outcomes and found different domains to be related to different outcomes (e.g., Curby et al., 2009; Gosse et al., 2013; Guo et al., 2010; McCormick et al., 2015; Xu et al., 2014;). A confirmatory factor analysis supported the fit of three domains over one- or two- factor structures (Hamre et al., 2013). In addition, I considered the validity of the two domains with lower internal consistency levels (*ES* and *CO*) for including items that represent teacher's use of strategies that are emotionally supportive, and strategies that promote routines and expectations. After an examination of the content of the items within each domain, I was satisfied that the items within each domain represent facets of the respective domain, even though they may have captured strategies that may not necessarily be highly related to one another. For each quality of delivery component, underlying dimensions were all positively correlated.

In addition to the rationale for examining the three highly correlated domains separately in order to avoid multi-collinearity and increased type 2 error (failing to reject a null hypothesis of no effect), the conceptual basis of the three separate domains that appeared to be valid, and published studies that support different effects of each of the domains on children's outcomes, supported this study's analytic approach to include the *ES*, *CO*, and *IS* domains in separate models.

Data Analysis

Data Preparation and Examination

Multiple datasets from the *MTP-M/S* study were examined in order to gather descriptions of data that could be used to evaluate the feasibility and appropriateness of this research design. Separate datasets contained child-level demographic, science assessments, teacher-level demographic, and fidelity measures data.

First, in all datasets, cases for participants in the treated classrooms were selected. Then when applicable, cases with science activity data were selected. For example, from the datasets with observation measure data, cases of science tapes from *Basic* and *Plus* groups were selected. Then, total scores were calculated for adherence and also for each of the three components of quality of delivery. Composite science scores were calculated from the two science measures (*LiS* and *EPS*), one for fall and one for spring as previously described in Composite Science Score. The interrater reliability statistics were calculated with the double coded scores for adherence and quality of delivery (see Procedures). From measurement datasets with item scores (science assessment, adherence, and quality of delivery), the internal consistency statistic for each measure was calculated to confirm their reliability (see Measures).

For the purpose of conducting the analysis for the research questions on interaction effects, the predictor variables of interest were transformed to standardized values. To create the interaction terms, a cross product of the standardized variable was calculated in combinations of two of the fidelity variables. This resulted in seven interaction terms—*Adherence X ES, Adherence X CO, Adherence X IS, Dosage X ES, Dosage X CO, Dosage X IS,* and *Adherence X Dosage*. Finally, all of the selected cases and their data for relevant variables were combined into one dataset in order to run the statistical analysis.

Covariates

Children's family socioeconomic status and mother's education level, teacher background, and type of professional support teachers received were included in the models as covariates.

Fall Science Score

A composite science score was calculated for both fall and spring. The fall science score was used as a covariate.

Treatment Condition

There may have been differences in teachers' behavior, other than adherence, dosage, and quality of delivery, that resulted from the differing levels of PD supports that the *Basic* and *Plus* teachers received. Therefore, to control for the effect of different levels of PD supports, it was included as a covariate. The level of professional support teachers received was dummy coded with *Basic* group as the reference group (*Basic* = 0, *Plus* = 1).

Family Characteristics

Families completed a demographic survey where they reported their income and number of people living in the home, which was then converted into an income-to-needs ratio using guidelines from the U.S. Census Bureau. Mothers also reported their highest level of education.

Teacher Characteristics

Teachers provided information about their years of pre-kindergarten work experience and the highest level of education they had obtained. Since a bachelor's degree was a requirement for teachers, a variable for highest level of education obtained was dummy coded with not having completed a master's degree as the reference category (no master's degree = 0, master's degree = 1).

Descriptive Analysis of Key Variables of Interest

In order to understand the values represented in the dataset, a descriptive analysis of mean and standard deviation was conducted for variables of interest. The distribution of children's science score appeared to have narrowed in the spring. A paired-sample t-test comparison showed that the difference between the fall and spring scores was significant, t (226) = -17.63, $p < .001$. Results showed science scores in the spring (*M* = 83.56, *SD* = 9.34) were higher than in the fall (*M* = 74.80, *SD* = 10.06). Adherence was a standardized score therefore had a mean value of 0 and standard deviation of 1 (range: - 1.77 – 0.92). On the *CLASS* measure's seven-point scale, the scores for *Emotional Support* and *Classroom Organization* were in the mid- to high-range (range: 4.25 – 5.79, *M* = 5.29, *SD* = 0.35; *CO*: range: 4.33 – 6.33, *M* = 5.38, *SD* = 0.57), while the scores for *Instructional Support* were in the low- to mid-range (range: 1.83 – 4, *M* = 2.97, *SD* = .50). On average, teachers submitted about 40 science tapes each with a wide range of dosage distribution (range: 2 – 66). Descriptive statistics for measures of interest and covariates are presented in Table 3.

Table 3

Covariates, Predictors, and Dependent Variable, Child (n = 328) and Teacher/Classroom (n = 31)

	N	Missing	*M*	*SD*
Child characteristics				
Fall science	306	22	74.16	9.98
Spring science [b]	252	76	83.32	9.45
Family income/needs ratio [a]	292	36	1.30	0.92
Maternal education [a]	314	14	4.25	1.57
Teacher characteristics				
Holds a master's [a]	31		0.52	0.51
Experience with preschoolers [a]	28	3	8.05	7.02
Classroom characteristics				
Plus [a]	14			
Basic [a]	17			
Measures				
Adherence (standardized score) [c]	29	2	0.00	1.00
Dosage [c]	31		40.13	21.07
Quality				
Emotional support [c]	29	2	5.29	0.35
Classroom organization [c]	29	2	5.38	0.57
Instructional support [c]	29	2	2.97	0.50

[a] Covariate
[b] Dependent Variable
[c] Predictors
[d] 1 = 8th grade or less, 2 = some high school but no diploma , 3 = high school diploma or equivalent, 4 = high school diploma or equivalent and technical certificate, 5 = some college but no degree, 6 = associate’s degree or two-year degree, 7 = bachelor’s degree, 8 = graduate degree.

Missing Data Imputation

The dataset was also screened for the presence of missing data at the case and variable levels in order to evaluate the quality of missing and available data. Teachers who dropped out of the study after submitting any data and their students, and children who were selected for assessments but missed one or more of the assessments were included in the analysis using full information maximum likelihood (FIML) method in *Mplus* (Muthén & Muthén, 2007). Full information maximum likelihood estimation analyses used all available data from each case when estimating parameters. When data was missing at random and there is multivariate normality, FIML estimation will be unbiased (Enders & Bandalos, 2001). Data is considered to be missing at random when the missingness is not related to the value of the missing data, value which cannot be observed. Compared to listwise or pairwise deletion methods that result in biased estimation and loss of statistical power, FIML is a superior method of dealing with missing data (Allison, 2002). Therefore, data for teachers and children were first examined as described below to ensure that missing data are missing at random before implementing FIML.

Missing Data for Teachers

Examination of the variables that were used for this study showed that 26 teachers had complete data. For teachers' demographic variables, missing data ranged between 0% to 12.9%. Five teachers (two *Basic* and three *Plus* teachers) dropped out of the study. Of these five teachers, the district pulled out four to participate in another study and one teacher pulled out due to a family tragedy.

Missing Data for Children

Of all selected children in the treatment groups (N = 328), 181 children (55.2%) had complete data for all variables of interest (dependent variable, independent variables, and covariates). Eighty-four children (25.6%) had missing data for one variable; 19 children (5.8%) had missing data for two variables; 17 children (5.2%) had missing data for three variables. Remaining 27 children (8.2%) had missing data for four or five variables. For child-level variables, missing data ranged between 4.2% to 23.4%.

Comparison of the students who participated for the whole study and those who did not participate for the whole study was conducted using an Analysis of Variance (ANOVA) test. The two groups of children did not differ significantly in science outcomes ($p > .05$). Their demographic profiles, classroom conditions, and assessment scores were not significantly different except for ethnicity. Children who did and did not participate in the whole study did not differ significantly in regard to age, gender, family income-to-needs ratio, and mother's highest level of education completed ($p > .05$). Children also did not differ on the levels of professional support provided to teachers (Plus or Basic), teacher's holding a master's degree, and teacher's years of experience in preschool. A significantly smaller proportion of children who did not participate in the whole study (i.e., missed one or more assessments) were Black/African-American ($M = .55$, $SD = .50$) compared to those who participated in the whole study ($M = .73$, $SD = .44$), $F(1, 320) = 10.14$, $p < .01$. A significantly larger proportion of children who did not participate in the whole study were White ($M = .33$, $SD = .47$) compared to those who participated in the whole study ($M = .20$, $SD = .40$), $F(1, 320) = 5.41$, $p < .05$. Table 4 compares children who participated in the whole study with those who missed one or more assessments.

Table 4

Mean Comparison of Students in the Whole Study and Missing Fall or Spring Assessments

	Whole study (*N* = 236)			Missed Assessment (*N* = 92)				
	Mean	SD	Range	Mean	SD	Range	F (df)	
Fall science	0.09	1.75	-4.33 – 4.21	-0.29	1.63	-6.05 – 3.80	2.65	(1, 304)
Spring science	0.03	1.74	-5.78 – 3.08	-0.33	2.20	-6.54 – 2.75	0.79	(1, 250)
Student age	4.62	0.33	2.92 – 5.71	4.61	0.31	3.82 – 5.06	0.07	(1, 299)
Student gender	0.54	0.50	[a]	0.52	0.50	[a]	0.06	(1, 319)
Student ethnicity, Black/African-American	0.73	0.44	[b]	0.55	0.50	[b]	10.14	(1, 320) **
Student ethnicity, White/Caucasian	.20	0.40	[c]	0.33	0.47	[c]	5.41	(1, 320) *
Maternal education	4.22	1.58	1 – 8 [d]	4.31	1.54	1 – 8 [d]	0.22	(1, 312)
Family income/needs ratio	1.30	0.96	0.06 – 4.70	1.30	0.83	0.05 – 3.50	0.00	(1, 290)
Teacher's years of experience, pre-k	7.85	7.35	2 – 32	7.84	5.51	2 – 32	0.00	(1, 293)
Teacher holds a master's degree	0.54	0.50	[e]	0.53	0.50	[e]	0.02	(1, 283)
Plus or Basic	1.58	0.50	[f]	1.50	0.50	[f]	1.56	(1, 326)

Note. * $p < .05$

** $p < .01$.

[a] 0 = male, 1 = female,

[b] 0 = not Black/African-American, 1 = Black/African-American

[c] 0 = not White/Caucasian, 1 = White/Caucasian

[d] 1 = 8th grade or less, 2 = some high school but no diploma , 3 = high school diploma or equivalent, 4 = high school diploma or and equivalent technical certificate, 5 = some college but no degree, 6 = associate’s degree or two-year degree, 7 = bachelor’s degree, 8 = graduate degree

[e] 0 = does not hold a master's degree, 1 = holds a master's degree

[f] 1 = Plus, 2 = Basic.

Data Screening for Multiple Regression Analysis

First, the dataset was assessed for extreme values, distribution of residuals and bivariate correlations between the independent variables. Regression models were examined for extreme values. Dependent variable values that are extremely far from the mean have leverage and can affect the regression coefficient estimates. Outliers have large residuals and are identified by their unusual dependent variable values given their predictor variable values. Residuals were examined for three separate models including all interaction terms. Each model included covariates, one of the quality of delivery variables (*Emotional Support*, *Classroom Organization*, or *Instructional Support*), adherence, dosage, and the cross products of quality of delivery variable, adherence, and dosage. Several cases were identified as exhibiting extreme values in leverage by using centered leverage values (cutoff > .155), having discrepancy by using distribution of studentized residuals (|cutoff| > 3), and having influence by using Cook's distance (cutoff > .022). Cutoffs were determined using the general rule of thumb. However, a judgment was made that the extreme values should be represented in this analysis because the data points did not have additional information with which they might be examined. After manually examining each case with extreme values and confirming that values for each variable of interest are within plausible range, the decision was made not to apply sweeping cutoffs. One intuition that supported keeping those cases was that for young children's academic performance, values that do not fit well along expected outcome value – very low or high scores – can reasonably be expected.

Next, the data were screened for assumptions of normality, equal distribution, and collinearity. Graphical examination of distributions of residuals using histogram, Q-Q and P-P plots showed the values met the assumptions of normality and equal distribution. Residuals had fairly symmetrical distribution with slightly negative skew (range: -0.06 – -0.11) and had leptokuric distribution with sharper peak and heavy tails (range: .73 – 1.08).

When assumption of homogeneity of variance is violated, the standard errors are biased and significance test cannot be trusted. Graphical examination of scatterplot of standardized residuals by standardized predicted values showed mild heteroscedasticity. However, mild heteroscedasticity is not likely to lead to serious bias (Allison, 1998).

Collinearity among variables was examined for all three full models. All independent variables in the three full models had acceptable tolerance (all tolerance > .1)—indicating that a small percent of variance cannot be accounted for by other independent variables in each model—and acceptable variance inflation factors (all VIF < 6) indicating that the predictors were correlated but not extreme enough to require correction based on the general rule of thumb.

In addition to the tolerance and variance inflation levels, the bivariate relationships of the independent variables were examined in order to ensure that the independent variables included in the models were not highly correlated. Several independent variables were weakly and significantly correlated to one another. Table 5 lists all correlations amongst the independent variables. As noted in published studies (e.g., Jamil et al., 2010), the three domains of quality of delivery (i.e., *Emotional Support, Classroom Organization*, and *Instructional Support*) had high and significant correlations. In the present study, *Emotional Support* and *Classroom Organization* (r = -.66, p < .01) and *Classroom Organization* and *Instructional Support* were highly correlated (r = -.56, p < .01), supporting the decision to examine the three domains of quality of delivery in separate models. For all other independent variables, although many significant correlations existed, the strength of the relationships were small or modest and thus did not raise concerns about multicollinearity.

Table 5

Bivariate Correlations between Independent Variables

	2	3	4	5	6	7	8	9	10
1. Study condition of classroom	$-.23^{**}$	$-.39^{**}$	.03	.07	$.17^{**}$	$.35^{**}$	$.17^{**}$	$.18^{**}$	-.05
2. Experience with preschoolers		.07	-.03	-.09	.07	-.07	$.24^{**}$	.49	$-.19^{**}$
3. Teacher holds master's			.03	-.06	-.02	$-.16^{**}$	-.07	-.12	$.26^{**}$
4. Income to need ratio				$.50^{**}$	.09	.09	.09	$.21^{**}$	-.10
5. Maternal education					.07	.10	.00	$.11^{*}$	-.04
6. Emotional support						$.66^{**}$	$.40^{**}$	$.38^{**}$	-.02
7. Classroom organization							$.56^{**}$	$.46^{**}$	.07
8. Instructional support								$.39^{**}$	$.17^{**}$
9. Adherence									.05
10. Dosage									--

Note. * $p < .05$, ** $p < .01$.

Data Analysis

For all of the hypotheses, multiple linear regression mixed models were analyzed. These models were estimated at the child level but were controlled both for the random effect of classroom (b_j) as well as the random effect of child (ε_{ij}). This allowed these models to appropriately account for within-classroom dependence, helping to ensure that the assumption of independence was met. This helped avoid the underestimation of standard errors, underestimated *p* values, and inflated type 1 error rate that can occur when analysis of nested data does not account for dependence of child level variables in the same group. The mixed models were estimated in *Mplus* (Muthén & Muthén, 2007) using analysis “Type = Complex,” identifying classroom as a cluster variable. In these models, the main and interaction effects were estimated as fixed factors.

All models included the covariates previously described: fall science score, *Basic* or *Plus* treatment condition (0 = *Basic*, 1 = *Plus*), teacher’s years of experience with preschool children (Pre-K Experience), whether the teacher holds a master’s degree (Teacher Education, 0 = No, 1 = Yes), the highest level of education completed by the female head of household (Maternal Education, 1 = less than 8^{th} grade to 8 = master’s degree), and family’s income-to-needs ratio (Family Income). These covariates were included to explain some of the variance in children’s science learning outcomes and thus reduced possible inflation of predictors’ effects. The fall science score used as a covariate in all models also served the purpose of allowing the interpretation of each predictor’s relationship with the change in science scores from fall to spring, i.e., gains in children’s science knowledge and skills.

Published studies on preschool children's learning suggest that each of the three quality of delivery components may be related to different types of outcomes (e.g., Instructional Support with academic and *Emotional Support* with social-behavioral outcomes) (e.g., Curby et al., 2009), and that the three quality of delivery components are highly correlated. Based on this and concerns about multicollinearity previously described, separate models that each included one of the three quality of delivery components with other predictors were analyzed as specified below.

The first research question "Does fidelity of curriculum implementation (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year?" was explored with one hypothesis and three models each containing one of three quality of delivery components.

R1H1: Adherence, dosage, and quality of delivery measured by teachers' use of teacher-child interaction strategies (*ES*, *CO*, and *IS*) each have a positive additive effect on children's gains in science learning outcomes.

- *Science Outcome Spring* = $\beta_0 + \beta_1$ *Science Outcome Fall* + β_2 *Plus* + β_3 *Pre-K Experience* + β_4 *Teacher Education* + β_5 *Maternal Education* + β_6 *Family Income* + ***β_7 Adherence + β_8 Dosage + β_9 ES*** + $b_j + \varepsilon_{ij}$
- *Science Outcome Spring* = $\beta_0 + \beta_1$ *Science Outcome Fall* + β_2 *Plus* + β_3 *Pre-K Experience* + β_4 *Teacher Education* + β_5 *Maternal Education* + β_6 *Family Income* + ***β_7 Adherence + β_8 Dosage + β_9 CO*** + $b_j + \varepsilon_{ij}$
- *Science Outcome Spring* = $\beta_0 + \beta_1$ *Science Outcome Fall* + β_2 *Plus* + β_3 *Pre-K Experience* + β_4 *Teacher Education* + β_5 *Maternal Education* + β_6 *Family Income* + ***β_7 Adherence + β_8 Dosage + β_9 IS*** + $b_j + \varepsilon_{ij}$

The models for the second research question—"Does the association of one fidelity component with children's gains in science learning outcomes depend on levels of another fidelity component?"—included the same covariates and main effects variables that have been specified above. For brevity, the omission of the covariates and main effect variables from the regression equations for the second research question are indicated by ellipsis.

Multiple regression analyses with interaction terms were performed in order to answer the second research question on whether fidelity components strengthen the association between other fidelity components and science outcomes. To test whether one of the fidelity components strengthens the association between another one of the fidelity components and science outcomes, three hypotheses were tested. For each of the types of quality of delivery, the three hypotheses regarding two-way interactions between three types of interactions were tested with the three interaction terms in the same model. In this way, three hypothesis questions below were addressed in each of the three models.

R2H1: The relation of adherence to gains in children's science learning outcomes is greater when teachers implement curriculum using higher quality of delivery (*ES*, *CO*, and *IS*).

R2H2: The relation of dosage to gains in children's science learning outcomes is greater when teachers implement curriculum using higher quality of delivery (*ES*, *CO*, and *IS*).

R2H3: The relation of adherence to gains in children's science learning outcomes is greater when teachers implement curriculum using higher dosage.

- *Science Outcome Spring =…* $\boldsymbol{\beta_9}$ ***ES*** **+** $\boldsymbol{\beta_{10}}$ ***Adherence x ES*** **+** $\boldsymbol{\beta_{11}}$ ***Dosage x ES*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage*** + $b_j + \varepsilon_{ij}$

- *Science Outcome Spring =…* $\boldsymbol{\beta_9}$ ***CO*** **+** $\boldsymbol{\beta_{10}}$ ***Adherence x CO*** **+** $\boldsymbol{\beta_{11}}$ ***Dosage x CO*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage*** + $b_j + \varepsilon_{ij}$
- *Science Outcome Spring =…* $\boldsymbol{\beta_9}$ ***IS*** **+** $\boldsymbol{\beta_{10}}$ ***Adherence x IS*** **+** $\boldsymbol{\beta_{11}}$ ***Dosage x IS*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage*** + $b_j + \varepsilon_{ij}$

(See Appendix E for mapping of research questions and hypotheses to analytic models.)

Power Analysis

The number of teachers and students needed to achieve adequate statistical power was derived in order to inform the necessary sample size required to detect a small effect size typically found in education research. This study drew from an existing sample and thus the sample size was already established. The purpose of the analysis was to determine the minimum effect size this study could detect given its sample size, in order to better understand its limitations. Power analyses were conducted with the following parameters and assumptions:

1. **Clustering levels**: The random assignment occurred at the school level. However, the primary level where children are interacting with one another is at the classroom level. Therefore, the amount of dependence among children's learning outcomes is likely to be greatest within classrooms. Thus, I choose the classroom level over the school level as the appropriate level for analysis.
2. **Cluster size**: An average of 10.58 children participated in each classroom.
3. **Intraclass correlation (ρ = .088) and cluster level covariates (**$\boldsymbol{R^2_{L2}}$ **= .074):** Intraclass correlation coefficient and cluster level covariates were calculated from the dataset using SPSS.

4. **Level of significance:** I used a two-tailed test and set the alpha level, the probability of making a type 1 error, to $p = .05$.
5. **Number of clusters ($N = 31$):** Students were exposed to the condition of the teacher's implementation in their respective classrooms.
6. **Power:** From the range of .7 to .8 suggested by Cohen (1977), I selected the higher .8 level.

Based on the parameters above, I used the *Optimal Design for Longitudinal and Multilevel Research - v1.55* (Raudenbush et al., 2005) to determine that given the cluster level student sample ($N = 10$) and number of clusters ($K = 31$) of the treatment group, the lowest effect size I would be able to detect is at the higher end of medium effect size ($\delta = 0.43$). In education research, the policy-relevant minimum effect size for a study design to detect has been in flux (0.10 to 0.20), but the typical level used in published studies is 0.20 (Bloom, 2008). Based on the parameters above, the number of clusters needed to detect minimum effect size of .20 with 80% power is 138, which is substantially larger than the 31 clusters observed in this study. Thus, the results of my study should be interpreted with the understanding that this study will be able to detect medium effect size but will miss smaller effect sizes.

Possible Bias: Validity and Quality

Bias due to validity and quality may limit interpretation and applicability of a study's findings. Given that the data for this study came from a previously conducted curriculum efficacy trial, it is possible to identify several possible sources of bias and threats to validity and quality and discuss how they were alleviated in the *MTP-M/S* study and in my study design.

The first consideration is selection bias in assignment to conditions, which was reduced through random assignment in the *MTP-M/S* study. In my study, selection bias was minimized by using all of the participants who had already been randomly assigned to the treatment condition, all of the randomly selected children in the treatment classrooms, and all of the scores from randomly selected video tapes submitted by teachers.

A second possibility is sample selection bias. Even though every pre-kindergarten teacher in the district had the opportunity to participate in the *MTP-M/S* study, one limitation of the design was each teacher's self-selected participation. Therefore, my study was composed of teachers who were more willing to participate in research, understanding that they may be asked to use new interventions and that their performance would be studied. Therefore, the random sample selection requirement for external validity, i.e., the generalizability to population of teachers working in Title IV schools, was not met. As such, this study is limited in its ability to make inferences about the general pre-kindergarten school environment from the findings. However, a true random sampling from the population is generally not possible in classroom-based research that requires voluntary participation. Therefore, whether using true random sampling from the general population of teachers in preschool classrooms is a realistic requirement for generalizability to larger population of teachers and classrooms is debatable.

Other aspects of the *MTP-M/S* study design allowed for high-quality data for the variables of interest. The duration of the study allowed for the assessments and observations to be taken from more than one time point over an academic year, thus adding to the validity of the measures. Pre- and post-assessments for science knowledge and skills were administered to children an average of 26 weeks apart (M = 26.47, SD = .79). Due to the elapsed time between administrations, children had diminished sensitization to the repeated measures. In addition, adherence and quality of delivery were observed from samples of videos obtained across the year, and thus the data reflected the average quality experienced by children over a time period. Thus, my study included data that reflected multiple implementation efforts by teachers over time, which was more useful for my research questions than observation at one or two time points. Also, the research team made efforts to support high quality in data collection through the training of direct assessors and video observation coders. For the observation measures, a portion of the selected videos was double coded so that interrater reliability could be determined.

Finally, a few limitations existed with possible errors in data collection and with the measures used to collect data. The data collected with paper forms were subject to human error when they were filled out by participants and assessors (e.g., teachers could have recorded the wrong birth date for one or more children in their classroom) and error could have been introduced in the transfer of information from paper to computer (e.g., interpreting a 1 as a 7 or mistyping numbers). In addition, using the tape count variable for the dosage measure poses some limitations. Some teachers may have submitted fewer tapes than the number of activities they implemented because they felt uncomfortable about the implementation quality in the recorded activities. Also, teachers who are less organized or less familiar with using the video recorder may have missed taping science activities due to forgetfulness or error in equipment operation. However, the tapes provide observable, verifiable evidence that the teacher on camera implemented the intended curriculum. For this reason, although an imperfect measure, a count of tapes submitted provides a superior measure of dosage than self-report with low compliance for the *MTP-M/S* study. Another source of data which presents some limitations is the children’s science assessment measure. As the *MTP-M/S* researchers pointed out, a validated science measure was not available at the time of the study. The science assessment and procedures were developed for the MTP-M/S study and it is possible that the science scores may not capture all science learning (Kinzie et al., 2014, p. 596).

In this chapter, I described the source of my data, curriculum, participants, data collection procedures, measurements, dependent and independent variables of interest, and covariates. Then, I recounted the steps I took to prepare and screen the data for analysis. I presented the hypotheses and accompanying models that served to answer my research questions. Finally, I examined the possible biases and limitations inherent in the study design. Next in Chapter 4, I present the findings from the multiple regression analyses, organized by research questions.

CHAPTER 4: RESULTS

Multiple Regression Mixed Model Results

A series of multiple linear regression analyses was performed to answer three research questions about the relationship between teachers' implementation of high-quality curriculum and children's science learning. First, regression models were used to test if children's gains in science knowledge and skills are predicted by teachers' fidelity to adherence, dosage, or quality of delivery. Then, regression models were used to test if components of teachers' fidelity interact to effect other component's associations with children's learning gains. A report of findings for each hypothesis and interpretation of significant effects are presented next. For the main effects models, β values indicate the amount of change in the standard deviation of outcome variable that can be expected from one standard deviation change in the independent variable. For the interaction terms, graphical representations of significant effects were used to interpret findings.

All models explained significant amount of variance. The main effects models explained approximately 53% to 54% of variance. The interaction models explained approximately 55% to 56% of variance.

In the following section, results of the analysis are interpreted and visually examined.

Research Question 1

Hypothesis 1: Main Effects of Fidelity Components

Table 6 shows the results from multivariate mixed regression analyses that examined the main effects of the implementation fidelity components on children's gains in science outcome.

All models accounted for an acceptable portion of the variance in science outcomes (all R^2's > .53). The research question was: Does curriculum implementation fidelity (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year? Among the quality of delivery components, *Instructional Support* had significant effect. No significant main effects were found for adherence (all p's > .05) and dosage (all p's > .05) in all of the models. The effect of *Instructional Support* was significant in the IS model (β = 0.077, p < .05) but quality of delivery was non-significant in the ES and CO models. In the IS model, increased *Instructional Support* was associated with greater change in science scores.

Table 6

Multivariate Main Effects Models Predicting Change in Science Scores

	ES model		CO model		IS model	
	β	S.E.	β	S.E.	β	S.E.
Fall science	0.693**	0.036	0.695**	0.037	0.698**	0.037
Plus	0.050	0.051	0.034	0.052	0.028	0.050
Holds a master's	0.049	0.057	0.043	0.059	0.016	0.060
Experience with preschoolers	-0.107	0.056	-0.105	0.060	-0.113	0.056
Family income/needs ratio	0.073	0.045	0.072	0.045	0.076	0.045
Maternal education	-0.020	0.063	-0.019	0.063	-0.022	0.063
Adherence	0.050	0.043	0.028	0.040	0.012	0.037
Dosage	-0.039	0.039	-0.043	0.050	0.013	0.038
Quality of delivery						
Emotional support	0.007	0.046				
Classroom organization			0.052	0.050		
Instructional support					0.077*	0.034
R^2	0.531**		0.531**		0.535**	

Note. ES = *Emotional Support*; CO = *Classroom Organization*; *IS* = Instructional Support
*p < .05, two-tailed. **p < .01, two-tailed.

Research Question 2

Table 7 shows the results from the multivariate mixed regression analyses with interaction terms that examined the moderation effects of fidelity components on the relation of another fidelity component to science outcomes. All models accounted for an acceptable portion of the variance in science outcomes (all R^2's > .56). The second research question asked: Does the association of one fidelity component with children's science outcomes depend on levels of another fidelity component?

Table 7

Two-Way Interaction Multiple Regression Models Predicting Change in Science Scores

	ES model		CO model		IS model	
	β	S.E.	*β*	S.E.	*β*	S.E.
Fall science	0.686**	0.041	0.700**	0.038	0.695**	0.035
Plus	0.041	0.045	0.028	0.054	0.051	0.049
Holds a master's	0.020	0.061	0.087*	0.042	0.111*	0.056
Experience with preschoolers	-0.078	0.056	-0.148*	0.067	-0.096	0.049
Family income/needs ratio	0.078	0.045	0.083	0.045	0.078	0.044
Maternal education	-0.030	0.061	-0.023	0.062	-0.029	0.064
Adherence	0.169*	0.076	0.148**	0.037	0.073	0.039
Dosage	0.039	0.054	-0.032	0.064	-0.045	0.049
Quality of delivery						
Emotional support (ES)	-0.126	0.065				
Classroom organization (CO)			0.021	0.047		
Instructional support (*IS*)					-0.006	0.050
Two-Way						
ES						
Adherence x ES	0.099	0.063				
Dosage x ES	0.191*	0.079				
Adherence x dosage	-0.114*	0.053				
CO						
Adherence x CO			0.162**	0.052		
Dosage x CO			0.055	0.062		
Adherence x dosage			-0.094**	0.036		
IS						
Adherence x IS					0.092**	0.028
Dosage x IS					0.108*	0.048
Adherence x dosage					0.026	0.027
R^2	0.566**		0.560**		0.545**	

Note. ES = *Emotional Support*; CO = *Classroom Organization*; IS = *Instructional Support.*
**p* < .05, two-tailed. ***p* < .01, two-tailed.

Hypothesis 1: Two-Way Interaction Effect of Adherence and Quality of Delivery

Relation of curricular adherence to gains in children's science knowledge and skills was significantly moderated by *Classroom Organization* ($\beta = 0.162$, $p < .01$) and *Instructional Support* ($\beta = 0.092$, $p < .01$), but not *Emotional Support* ($\beta = 0.099$, $p > .05$). Adherence appeared to be positively associated with science gains when *Classroom Organization* was high, whereas adherence did not appear to be associated with science gains when *Classroom Organization* was low (Figure 6). A similar pattern was found in the model with *Instructional Support*. Adherence appeared to be positively associated with science gains when *Instructional Support* was high, whereas when *Instructional Support* was low, adherence did not appear to be associated with science gains (Figure 7).

Figure 10

Interaction Effect between Adherence and Classroom Organization

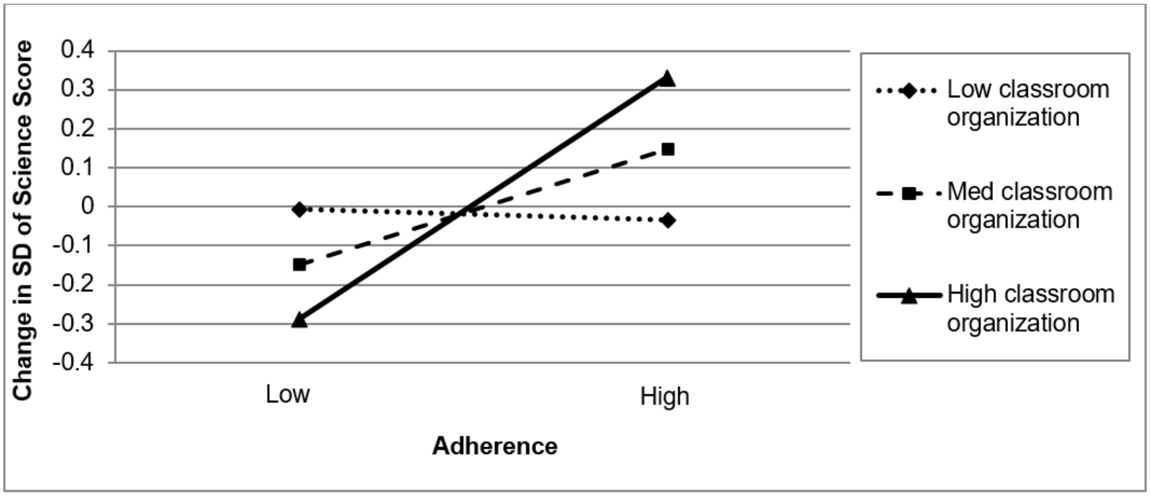
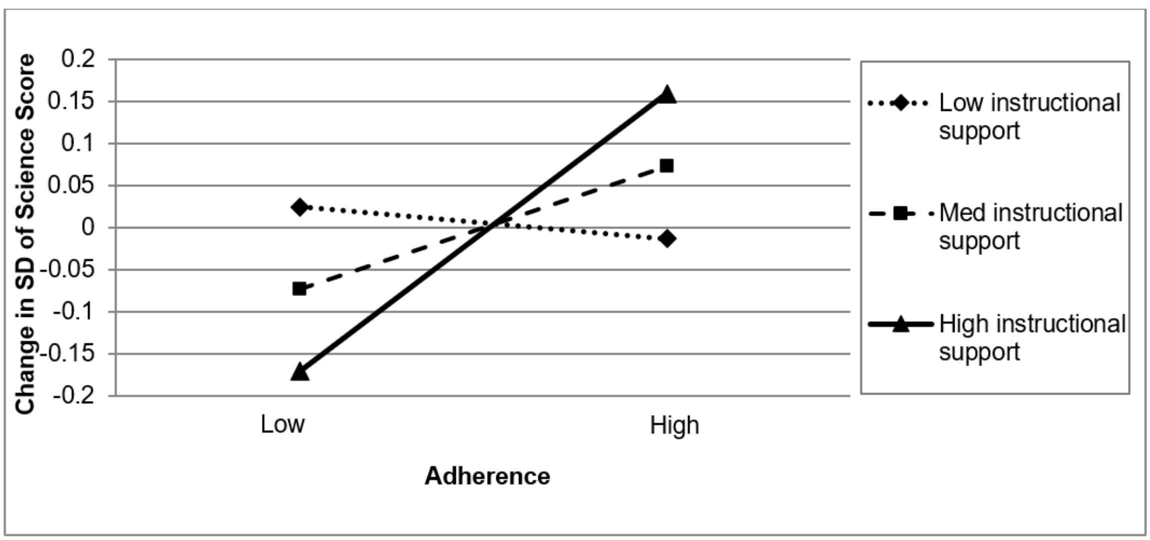


Figure 11

Interaction Effect between Adherence and Instructional Support



Hypothesis 2: Two-Way Interaction Effect of Dosage and Quality of Delivery

Relation of dosage to gains in children's science outcomes was significantly moderated by *Emotional Support* (β = 0.191, p < .05) and *Instructional Support* (β = 0.108, p < .05), but not

Classroom Organization ($\beta = 0.055$, $p > .05$). Dosage appeared to be positively associated with science gains when *Emotional Support* was high, whereas dosage was negatively associated with science gains when *Emotional Support* was low (Figure 8). Similarly, dosage appeared to be positively associated with science gains when *Instructional Support* was high, whereas dosage appeared to be negatively associated with science gains when *Instructional Support* was low (Figure 9).

Figure 12

Interaction Effect between Dosage and Emotional Support

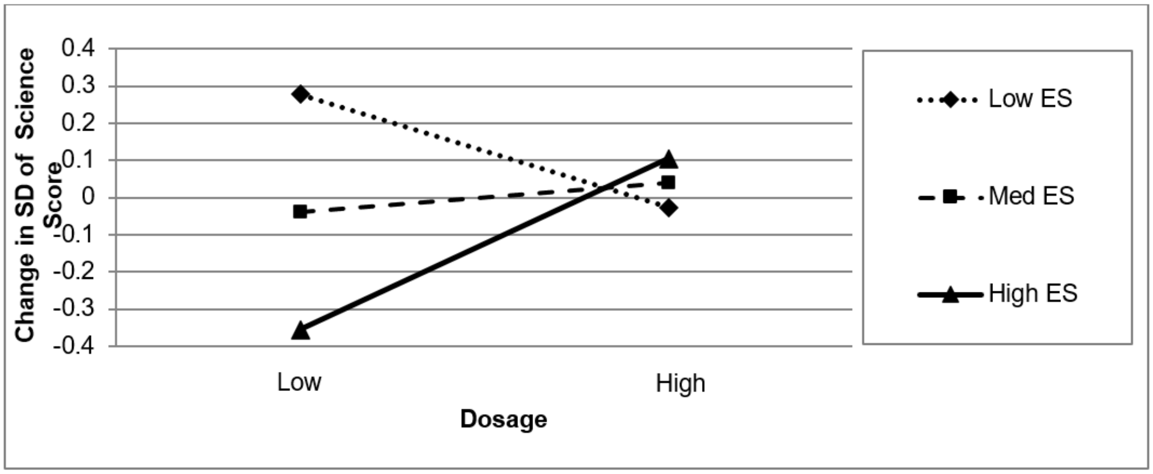
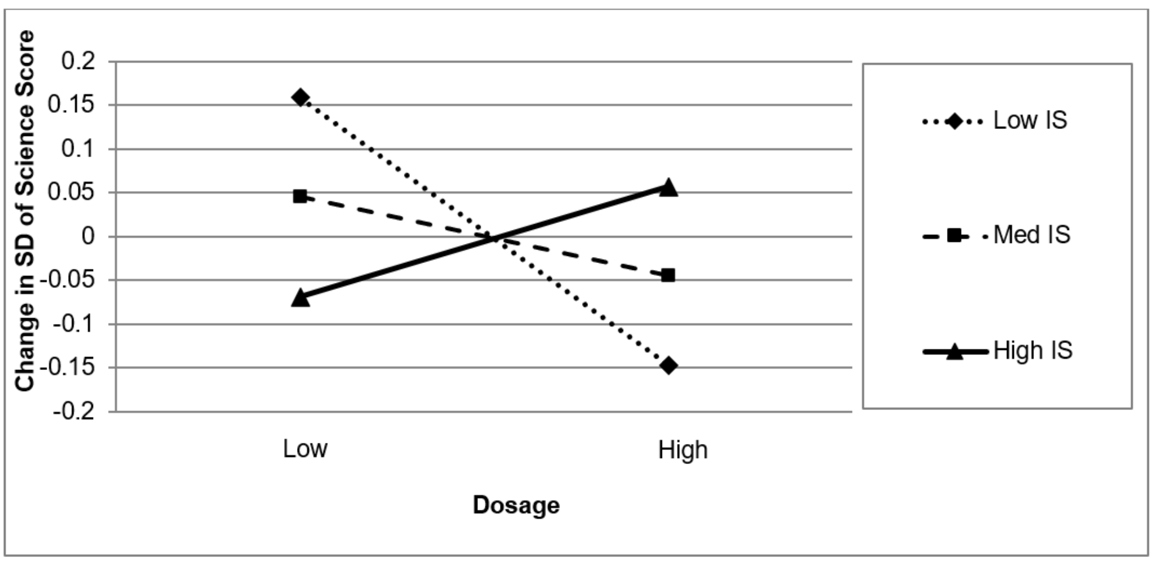


Figure 13

Interaction Effect between Dosage and Instructional Support



Hypothesis 3: Two

Two-Way Interaction Effect of Adherence and Dosage

The relation of adherence to children's gains in science knowledge and skills was significantly moderated by dosage in the model with *Emotional Support* (β = -0.114, p < .05) and the model with *Classroom Organization* (β = -0.094, p < .01), but not the model with *Instructional Support* (β = 0.026, p > .05). In both models, adherence did not appear to be associated with science gains when dosage was high, whereas adherence appeared to be positively associated with science gains when dosage was low. See Figure 10 for *Emotional Support* model results and Figure 11 for *Classroom Support* model results.

Figure 14

Interaction Effect between Adherence and Dosage, Emotional Support Model

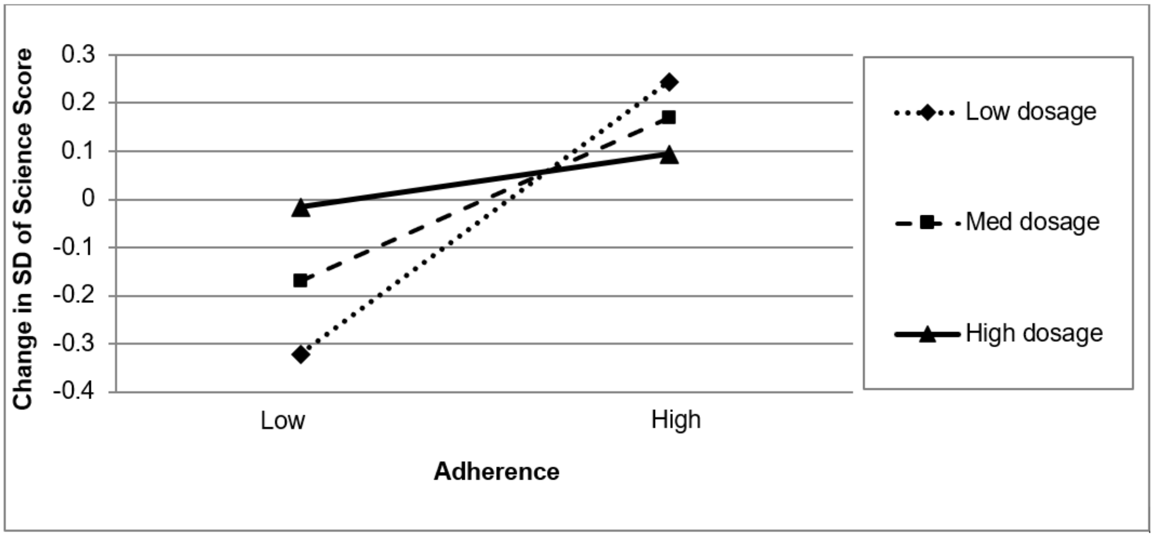
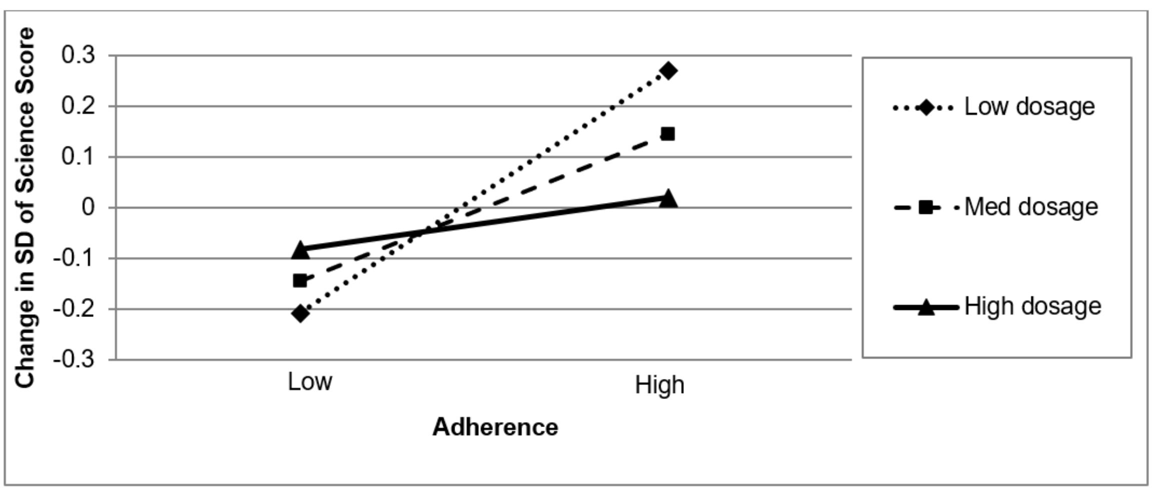


Figure 15

Interaction Effect between Adherence and Dosage Classroom Organization Model



Summary

Multivariate regression analysis was run to test four hypotheses to answer two research questions, one about main effects and the other about the interaction effects of fidelity components on children's gains in science scores. In the main effects analyses, only one sub-component of quality of delivery had significant effect; *Instructional Support* positively and significantly predicted children's gains in science outcomes. The two-way interaction analyses showed several indications of significant moderation effects. First, the association between adherence and children's gains in science scores was significantly moderated by *Classroom Organization* and *Instructional Support*. Adherence appeared to be positively associated with science gains when *Classroom Organization* and *Instructional Support* were high, but not associated with science gains when the quality of delivery sub-components were low. Second, relation of dosage to children's science outcomes was significantly moderated by *Emotional Support* and *Instructional Support*. Dosage appeared to be positively associated with science gains when either of the two quality of delivery sub-components was high, but negatively associated when the sub-components were low. Third, dosage significantly moderated the relation of adherence of children's gains in science outcomes in the *Emotional Support* and *Classroom Organization* models. In the two models, when the dosage was high, adherence did not appear to be associated with science gains, but when dosage was low, adherence appeared to be positively associated with science gains. In the next and final chapter, I explore the possible implications of these results.

CHAPTER 5: DISCUSSION

Increasingly, preschool teachers are expected to teach science. Research suggests that high-quality curricula implemented with fidelity can increase children's learning (O'Donnell, 2008). Thus, research has focused on supporting teachers to implement curricula with high fidelity. However, not enough is known about which key ingredients of curriculum implementation fidelity actually account for children's gains in learning (Durlak, 2010).

This study contributes to the body of knowledge on curriculum implementation fidelity by examining which aspects of fidelity are related to children's science outcomes, and how those fidelity components might interact to strengthen the relationship of other fidelity components to children's outcomes. Using the most often examined aspects of teachers' curriculum intervention fidelity—adherence, dosage, and quality of delivery—this study showed that fidelity components may not all be individually associated with children's science outcomes, and that *Instructional Support* and *Classroom Organization* may strengthen the contribution of other components to children's outcomes.

Summary of Findings

The individual components of fidelity generally did not predict children's learning outcomes. Only one sub-component of quality of delivery, *Instructional Support*, predicted children's learning outcomes. In addition, results showed two-way interaction effects between fidelity components. The relationship between adherence and children's learning outcomes was strengthened by higher quality of delivery, specifically *Instructional Support* and *Classroom*

Organization. When levels of quality of delivery were low, the positive association between adherence and change in children's learning outcomes disappeared. Similarly, dosage was positively associated with children's gains in learning when quality of delivery—*Emotional Support* and *Instructional Support*—was high. Notably, dosage was negatively associated with science gains when *Instructional Support* was low. Even more surprising, in the models with *Emotional Support* and *Classroom Organization*, where adherence was positively associated with science gains when dosage was low, adherence was not associated with science gains when dosage was high.

Interpretation

Among the components of fidelity examined in this study, only the *Instructional Support* was a significant predictor of children's science learning gains. This finding aligns with studies that found positive effects of the quality of teachers' interactions, specifically *Instructional Support* strategies (e.g., providing feedback and encouraging reasoning) on children's academic outcomes (e.g., Burchinal et al., 2010; Mashburn et al., 2008).

Analysis of interaction effects indicated that teachers' adherence to the strategies prescribed in the curriculum was positively related to children's learning gains when teachers used high quality of delivery. The number of science lessons teachers used with children (dosage) was also positively related to children's learning gains when teachers used high quality of delivery. Examining these results together, it appeared that quality of delivery strengthened the relationship between gains in children's science learning outcomes and teachers' adherence to prescribed guidance in the lessons, as well as dosage—the number of science lessons teachers implemented. When teachers used high quality teacher-child interaction strategies, the relationship between adherence and gains in children’s learning outcomes was positive, but when they used lower quality teacher-child interaction strategies, their adherence to lesson components was not related to children’s outcomes. When teachers used high quality teacher-child interaction strategies, the relationship between the number of science lessons that teachers implemented and gains in children’s learning outcomes was positive, but when they used lower quality teacher-child interaction strategies, the number of science lessons teachers implemented had a negative relationship with children’s learning gains.

The unexpected direction of the relationship between adherence and learning gains with different levels of dosage is a counterintuitive finding. Whereas it is typically assumed that high adherence to science lessons and high dosage (optimally a planned number of lessons incorporated into a high-quality curricula wherein no lesson is superfluous) are necessary to help children learn science, this result indicates the relationship between these two components and science learning outcomes may not be this straightforward. Considered together with the moderation effects found between adherence and quality, and between dosage and quality, this result could be an indication that the two-way interaction model does not sufficiently describe the relationship among fidelity components in their association with children’s learning outcomes. Since quality of delivery appears to moderate the effects of both adherence and dosage on children’s science learning outcomes, perhaps a three-way interaction analysis may better address the relationships. A three-way interaction model that may better explain the relationship might be one wherein the interaction effects of dosage on the association between adherence and children’s learning outcomes vary by levels of quality of delivery.

Implications

The outcomes of this study suggest that the goal of improving children's science learning outcomes through teachers' implementation of high-quality curricula is best supported when teachers implement the prescriptive elements of curriculum at higher dosage and engage in high-quality interactions. For instance, when teachers taught with high adherence or when teachers implemented high number of science lessons, higher quality of delivery strengthened their relationship with children's science learning gains. Without high quality of delivery, neither adherence to script and guidance, nor implementing many science lessons, were related to children's gains in science learning. In fact, implementing science lessons may even be a detriment to children's science learning outcomes when done with low quality teacher-child interaction.

Among the quality of delivery sub-components, *Instructional Support* persistently appeared to be an important moderator for both adherence and dosage. However, it should not be overlooked that *Emotional Support* and *Classroom Organization* were also significant moderators of adherence and dosage, respectively. The moderating role of the three types of teacher-child interaction strategies in this study supports the argument made for these developmentally supportive practices as having a role beyond limited types of outcomes—i.e., emotionally supportive practices with socio-emotional outcomes and cognitively supportive practices with academic outcomes (Downer et al., 2010). Whatever their individual contribution to children's science outcomes, developmentally appropriate teacher-child interaction strategies may have implications for both strength and direction of the relationship between children's science outcomes and content-specific strategies captured in the curricula, as well as the number of lessons implemented.

Thus, one way that science curriculum researchers and designers can improve the connection between the teachers' fidelity of curriculum implementation and the children's science outcomes would be to not only focus on adherence and dosage, but also to help teachers implement curricula while tending to children's developmental needs. As lessons in curricula are limited in how much guidance they can provide for teacher-child interaction strategies that are specific to the context of each classroom, a two-pronged approach may be appropriate. One part would be to ensure that high adherence and dosage can be achieved by real teachers in real classrooms with high quality of delivery through iterative curriculum design, development, and evaluation process (e.g., Kinzie et al., 2015).

The second part would be to include professional development and supports that help teachers use science-specific instruction in concert with high-quality interactions with children. This may be achieved by including this objective in the pre-service teacher education curricula or in workshops for in-service teachers. Preparing teachers to use the curriculum they will be teaching in their classrooms can have a positive effect on students' learning (Boyd et al., 2009). Although studies on professional development for preschool teachers' science instruction are few, existing studies on professional development for preschool teachers give some indication that high quality professional development support can improve teachers' classroom practices, namely improve quality of teacher-child interaction strategies (e.g., Pianta, Mashburn et al., 2008; Zan et al., 2014) and increase instances of science instruction (Piasta, Logan et al., 2015). Approximately 50 hours of professional development has been shown to result in high fidelity of curriculum implementation (Clements & Samara, 2008). Sustained professional development programs that allow for in-classroom practice and self-reflection, and provide consultation and feedback have been shown to help teachers improve the quality of teacher-child interaction strategies (Pianta, Logan et al., 2008; Zan et al., 2014). Combining curriculum implementation with professional development supports may be beneficial to teachers' instructional practices (Domitrovich et al., 2009).

Limitations

This study found that not all individual components of fidelity significantly predicted children's science learning gains. This should not be interpreted to mean that no relationship exists between those components and children's science learning gains. This only suggests that this particular study was not able to provide evidence to the contrary.

As previously discussed through power analysis, the sample size of this study limited the its ability to detect significant relationships to medium effect size. A larger sample size may allow for detection of relationships between each of the fidelity components and children's science gains. Another possible explanation for why each fidelity component was not significantly related to children's learning gains in this study is that the fidelity components individually may not have provided sufficient condition for children's science gains, or that the measures may not have been sensitive enough to capture the necessary conditions. For example, while quality of delivery has been shown to be related to children's academic outcomes in other studies (e.g., Curby et al., 2013), those significant effects were reported mostly for language and some for mathematics outcomes, not science. But conceptualization of how quality of delivery functions in preschool children's science learning have, necessarily, relied on those existing studies. This study contributes to this field and line of investigation specific to preschool science teaching.

Some limitations in measures may have also hindered this study's capacity to detect significant differences. As stated by the *MTP-M/S* researchers, the science assessment measure may not have been sensitive enough to capture the range of science learning in preschool settings (Kinzie et al., 2014). In addition, the counting of lessons may have been an insufficient representation of dosage; perhaps frequency, duration, and/or intervals should have also been examined.

In addition, the adherence instrument may not have captured all of the core science content specific strategies that are important for children's science learning gains. Together with the iterative process used in the curricular design, successful and unsuccessful implementation with diverse groups and settings should be studied to iteratively identify the core components (Blasé et al.,2012). Compared to mathematics (e.g., Learning and Teaching Early Math: The Learning Trajectories Approach by Clements, 2009), preschool science does not yet have well-developed and specific strategies and learning trajectories. Delineating and describing children's science learning trajectories with teaching strategies for building children's understanding of science content and science processes would contribute to the understanding of core science-specific strategies that teachers should use in implementing curriculum.

Also, the information about teachers' classroom practice and children's science learning experience depended on teachers' voluntarily recording videos of every instance of their implementation and sending every recorded instruction to the research team. This leaves room for incompleteness in data. For example, it is possible that recordings of weaker instructional quality were withheld by teachers, or that high-quality instructions were not recorded by teachers who simply forgot to record the event. Also, data on children's experience with science learning outside of the curriculum implementation time was not collected. If the teachers engaged in other types of science activities outside of the use of curricula, or if some children were involved in learning science outside of the classroom, this study could not account for these possible contributors to children's science learning.

Another limitation of this study was the study context, settings, and participants. The results may generalize to preschool classrooms of similar demographics of children and teachers. But since the preschools volunteered for the study, its findings have limited implications for a broader, more diverse sample of preschools wherein schools are given mandated curricula and teachers implement curricula without much choice. Also, about half of the teachers in this study had a master's degree. This level of education may not represent the norm of preschool teachers as the profession does not require this level of education for entry. State licensing rules vary; some states require a high school diploma and teaching certification, while others require a Bachelor's degree and teaching certification or license. A larger, more diverse representation of participants from different geographic areas may provide a better representation of typical preschool teachers' practice.

Although teaching is a complex phenomenon, conceptual simplification was necessary to explain and test which components in teachers' fidelity of curriculum implementation are important for children's science learning outcomes and how those practices may interact. This study put forth a purposefully defined conceptualization of teachers' curriculum implementation fidelity in order to build on widely studied conceptualizations in curriculum fidelity research, and conceptualizations of preschool teachers' instructional strategies. Therefore, the conceptualization used in this study may have excluded facets of fidelity and teaching strategies that have received little or no attention from the preschool curriculum research community but may be important for children's learning.

Future Research

Measuring and reporting fidelity would improve confidence in efficacy and effectiveness studies' outcomes (NRC, 2004). However, without a clear and shared standard against which the "research-based" curricula could be evaluated, the claim to be "research-based" is not useful (Clements, 2007). Measuring and reporting disparate fidelity figures without clear insight about how the fidelity components boost children's learning has limited meaning for the quality of the curricula, quality of instruction, and quality of children's instructional experience. More studies that provide clearly defined, contextualized conceptualization of fidelity components and examine the relationship to and amongst fidelity components could help firstly to improve curricula and secondly to improve the supports teachers need in order to tend to those components. Increased investigation into curriculum implementation fidelity could help establish some shared standards for how researchers evaluate curricula and fidelity.

Another approach to further develop our understanding of fidelity of curriculum implementation and its role in improving children's learning would be to examine its relationship with teachers' knowledge, skills, and attitude. Expertise is domain-specific and requires practice (Johnson et al., 1981). Curricula are tools which contain this domain-specific knowledge and the skills researchers and curriculum designers have theorized to be important for teachers to enact. A better understanding of how teachers' existing knowledge, skills, and attitude contribute to and are influenced by curriculum implementation and curriculum implementation fidelity may contribute to changes in teachers as well as in children's learning outcomes.

For example, teacher efficacy may contribute to children's learning and may interact with fidelity components. Researchers have theorized that teachers' self-efficacy is directly and indirectly related to students' academic learning outcomes (Goddard et al., 2000; Guo et al., 2012; Zee et al., 2016). Although not a curriculum implementation study, one study found that preschool teachers' self-efficacy interacted with teacher-interaction quality's association with children's literacy outcomes (Guo et al., 2010). These types of research efforts would contribute to closing the conceptual links among the curriculum implementation, teaching strategies, and teacher skills, knowledge, and attitude.

Conclusions

In order to understand the phenomenon of teachers' curriculum implementation and its relationship with learning outcomes, it is necessary to have clearly defined fidelity components. While a universal measure of fidelity of curriculum implementation may not be possible or appropriate, a foundational framework by which fidelity could be operationalized in the context of preschool classrooms could help researchers significantly. Such a framework would enable researchers to compare and interpret studies to better understand fidelity, teachers' practice when using curriculum, and curriculum design (O'Donnell, 2008). The clarity of these components can determine the replicability of teachers' curriculum implementation and children's learning outcomes across diverse schools and classrooms.

In summary, this study built its conceptualization on frequently examined components of fidelity in curriculum research. Through quantitative analysis, this study examined relationships between the components of fidelity of curriculum implementation—adherence, dosage, and quality of delivery—and children's science learning outcomes. It also examined the relationship amongst adherence, dosage, and quality of delivery in their relationship with children's science learning outcomes.

This study contributes to the currently limited body of research on teachers' fidelity of implementation of preschool science curriculum. The findings from this study lend support to the idea that teachers' ability to tend to all three components of fidelity matters for children's science outcomes. There is more to curriculum implementation fidelity than closely following lesson guidelines/instructions and using the curricula as many times as prescribed. Rather, teachers' use of high-quality teacher-child interaction strategies and leading children in science learning activities with lesson plans as prescribed, does not have a zero-sum relationship in benefiting children. This study showed that science learning for young children may be best supported by the presence of teacher strategies that tend to the content-learning objectives and developmental objectives. Based on the findings from this study, curriculum researchers and designers should consider whether their curriculum design and implementation support plans address two aspects: how teachers can closely follow their prescription, in concert with high-quality teacher-child interactions.

REFERENCES

Achieve. (2019). *The state of state science education policy*. *Achieve's 2018 science policy survey*. https://www.achieve.org/2018-science-policy-survey

Achieve (2013). *Closing the expectations gap*: *Annual report on the alignment of state K-12 policies and practice with the demands of college and careers*. https://www.achieve.org/2018-science-policy-survey

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

Allison, P. D. (2002). *Missing data*. Sage.

Arnett, J. (1986). *Caregivers in day care centers: Does training matter?* (UMI No. 8705689) [Doctoral dissertation, University of Virginia]. ProQuest Dissertations Publishing

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Ball, D. L., & Cohen, D. K. (1996). Reform by the book--What is or might be--the role of curriculum materials in teacher learning and instructional reform? *Educational Research*, *25*(9), 6–8. http://www.jstor.org/stable/1177151

Beauchamp, G. A. (1982). Curriculum theory: Meaning, development, and use. *Theory Into Practice*, *21*(1), 23-27. https://doi.org/10.1080/00405848209542976

Blase, K., & Fixsen, D. L. (2013). Core intervention components: identifying and operationalizing. *ASPE Research Brief*, (February), 21. https://doi.org/10.1177/1049731509335549

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289-328. https://doi.org/10.1080/19345740802400072

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, 44(3), 701–731. https://doi.org/10.3102/0002831207306743

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Education Evaluation and Policy Analysis*, *31*(4), pp. 416-440. https://doi.org/10.3102/0162373709353129

Bronfenbrenner, U. (2005). Ecological models of human development. In M. Gauvain & M. Cole (Eds.), *Readings on the Development of Children*, (4th ed., pp. 3-8). Worth.

Bronfenbrenner, U., & Evans, G. W. (2000). Developmental science in the 21st century: Emerging questions, theoretical models, research designs and empirical findings. *Social Development*, *9*(1), 115-125. https://www.doi.org/10.1111/1467-9507.00114

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. https://www.doi.org/10.1016/j.ecresq.2009.10.004

Bustamante, A. S., White, L. J., & Greenfield, D. B. (2018). Approaches to learning and science education in Head Start: Examining bidirectionality. *Early Childhood Research Quarterly*, *44*, 34–42. https://doi.org/10.1016/j.ecresq.2018.02.013

California Department of Education (2008). *California preschool learning foundations*. California Department of Education. https://www.cde.ca.gov/sp/cd/re/psfoundations.asp

Century, J., & Cassata, A. (2016). Implementation Research: Finding Common Ground on What, How, Why, Where, and Who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. https://doi.org/10.1177/1098214010366173

Clements, D. H. (2007). Curriculum research: toward a framework for “research-based curricula. *Journal for Research in Mathematics Education*, *38*(4), 35–70. https://doi.org/10.2307/30034927

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494. https://doi.org/10.3102/0002831207312908

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press.

Curby, T. W., Brock, L. L., & Hamre, B. K. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Education & Development*, *24*(3), 292–309. https://doi.org/10.1080/10409289.2012.665760

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., & Barbarin, O. (2009). The relations of observed pre-k classroom quality profiles to children's achievement and social competence. *Early Education & Development*, *20*(2), 346–372. https://www.doi.org/10.1080/10409280802581284

Dane, A. V, & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://www.doi.org/10.1016/S0272-7358(97)00043-3

Darrow, C. L. (2013). The effectiveness and precision of intervention fidelity measures in preschool intervention research. *Early Education & Development*, *24*(8), 1137–1160. https://www.doi.org/10.1080/10409289.2013.765786

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, *46*(2), 567–597. https://doi.org/10.3102/0002831208328089

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & Sanford Derousie, R. M. (2010). Implementation quality: lessons learned in the context of the head start REDI trial. *Early Childhood Research Quarterly*, *25*(3), 284–298. https://www.doi.org/10.1016/j.ecresq.2010.04.001

Domitrovich, C. E., & Greenberg, M. T. (2004). Introduction to the special issue: preventive interventions with young children: Building on the foundation of early intervention programs. *Early Education and Development*, *9289*, 37–41. https://www.doi.org/10.1207/s15566935eed1504_1

Downer, J. (2013). Applying lessons learned from evaluations of model early care and education programs to preparation for effective implementation at scale. In T. Halle, A. Metz, & I. M. Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 157-169). Brookes.

Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher–child interactions in the classroom: toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education and Development*, *21*(5), 699–723. https://doi.org/10.1080/10409289.2010.497453

Duncan, R. (1995). Piaget and Vygotsky revisited: dialogue or assimilation? *Developmental Review*, *15*(4), 458–472. https://www.doi.org/10.1006/drev.1995.1019

Durán, L. K., Gorman, B. K., Kohlmeier, T., & Callard, C. (2016). The feasibility and usability of the read it again dual language and literacy curriculum. *Early Childhood Education Journal*, *44*(5), 453–461. https://doi.org/10.1007/s10643-015-0729-y

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly*, *25*(3), 348–357. https://www.doi.org/10.1016/j.ecresq.2010.03.003

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256. https://www.doi.org/10.1093/her/18.2.237

Early, D., Barbarin, O., Bryant, B., Burchinal, M., Chang, F., Clifford, R., Crawford, G., Weaver, W., Howes, C., Ritchie, S., Kraft-Sayre, M., Pianta, R., & Barnett, W. S. (2005). *Pre-kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten and state-wide early educational programs (SWEEP) study*. National Institute for Early Education Research, Frank Porter Graham Child Development Institute. http://fpg.unc.edu/node/4654

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457. https://doi.org/10.1037/1082-989X.6.4.352

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly*, *19*(1), 138–149. https:/doi.org/10.1016/j.ecresq.2004.01.004

Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change, 1*(1), 5-27. https://doi.org/10.1023/A:1010068703786

Futrell, M. H. (1987). Public schools and four-year-olds: A teacher's view. *American Psychologist*, *42*(3), 251–253. https://doi.org/10.1037/0003-066X.42.3.251

Fixen, D. L.; Naoom, S. F.; Blasé, K. A.; Friedman, R. M.; Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). University of South Florida Louis de la Parte Florida Mental Health Institute.

Gallo-Fox, J., & Cuccuini-Harmon, C. M. (2018). The non-tested years: Policy's impact on early childhood curriculum. *Educational Forum*, *82*(4), 475–490. https://doi.org/10.1080/00131725.2018.1461524

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, *31*(1), 79–88. https://doi.org/10.1016/j.cpr.2010.09.007

Gibson, J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Lawrence Erlbaum Associates.

Glassman, M., & Zan, B. (1995). Moral activity and domain theory: An alternative interpretation of research with young children. *Developmental Review*, *15*(4), 434-457. https://doi.org/10.1006/drev.1995.1018

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2011). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness*, *4*, 25–52. https://doi.org/10.1080/19345747.2010.487927

Gosse, C. S., McGinty, A. S., Mashburn, A. J., Hoffman, L. M., & Pianta, R. C. (2013). The role of relational and instructional classroom supports in the language development of at-risk preschoolers. *Early Education and Development*, *25*(1), 110–133. https://doi.org/10.1080/10409289.2013.778567

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research, and practice. *Promotion of mental health and prevention of mental and behavioral disorders* (Volume 3). Substance Abuse and Mental Health Services Administration.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science readiness. *Early Education & Development*, *20*(2), 238–264. https:/doi.org/10.1080/10409280802595441

Guo, Y., Dynia, J. M., Logan, J. A. R., Justice, L. M., Breit-Smith, A., & Kaderavek, J. N. (2016). Fidelity of implementation for an early-literacy intervention: Dimensionality and contribution to children’s intervention outcomes. *Early Childhood Research Quarterly*, *37,* 165–174. https://doi.org/10.1016/j.ecresq.2016.06.001

Guo, Y., McDonald Connor, C., Yang, Y., Roehring, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders’ literacy outcomes. Elementary School Journal, 113, 3–24. doi:10.1086/665816

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. Teaching and Teacher Education, 26(4), 1094–1103. https://doi.org/10.1016/j.tate.2009.11.005

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities:

Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*(3), 329–347. https:/doi.org/10.1016/j.ecresq.2009.07.002

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49–84). Brookes.

Hamre, B. K., Pianta, R. C., Downer, J. T., Decoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Capella, E., Atkins, A., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness. *The Elementary School Journal*, *113*(4), 461–487. https://doi.org/10.1086/669616

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher-child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, *36*, 561–571. https://doi.org/10.1016/j.ecresq.2015.09.005

Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014). Measuring the interactive skills of caregivers in child care centers: Development and validation of the caregiver interaction profile scales. *Early Education and Development*, *25*(5), 770–790. https:doi.org/10.1080/10409289.2014.840482

Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, *25*(2), 235–250. https:/doi.org/10.1016/j.ecresq.2009.11.003

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. Early Childhood Research Quarterly, 23, 27–50. doi:10.1016/j.ecresq.2007.05.002

Hulleman, C. S., Rimm-Kaufman, S. E., & Abry, T. (2013). Innovative methodologies to explore implementation: Whole-part-whole-construct validity, measurement, and analytical issues for intervention fidelity assessment in education research. In T. Halle, A. Metz, & I. M. Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 65-93). Brookes.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Jamil, B. F. M., Downer, J. T., & Pianta, R. C. (2012). Association of pre-service teachers' performance, personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education Quarterly*, *39*(4) p119-138. https://eric.ed.gov/?id=EJ1001446

John, M. S., Sibuma, B., Wunnava, S., Anggoro, F., & Dubosarsky, M. (2018). An iterative participatory approach to developing an early childhood problem-based STEM curriculum. *European Journal of STEM Education*, *3*(3), 1-12. https://doi.org/10.20897/ejsteme/3867

Johnson, P. E., Zualkernan, I., & Garber, S. (1987). Specification of expertise. *International Journal of Man-Machine Studies*, *26*(2), 161–181. https://doi.org/10.1016/S0020-7373(87)80089-5

Justice, L. M., Mashburn, A., Hamre, B., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*(1), 51–68. https://doi.org/10.1016/j.ecresq.2007.09.004

Kagan, S. L. (1989). Early care and education: reflecting on options and opportunities. *The Phi Delta Kappan*, *71*(2), 104–106. http://www.jstor.org/stable/20404081

Kansas State Department of Education. (2014). *Kansas early learning standards.* Kansas State Department of Education. http://www.sciencedirect.com/science/article/B6VRT-4GBM2KH-3/2/c1604066d9d356475af58015319d869b

Katz, L.G. (2003). Program content and implementation. In US Department of Education (Ed.), USA *OECD thematic review of education and care policy*. Office of Educational Research and Improvement, Organisation for Economic Co-operation and Development http://www.oecd.org/unitedstates/27856788.pdf

Kinzie, M. B., Whittaker, J. V., McGuire, P., Lee, Y., & Kilday, C. (2015). Research on curricular development for pre-kindergarten mathematics and science. *Teachers College Record*, *117*(7), 1–40. https://eric.ed.gov/?id=EJ1059955

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children's mathematics and science learning. *Early Childhood Research Quarterly*, *29*(4), 586–599. https://doi.org/10.1016/j.ecresq.2014.06.007

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*(3), 155-178. https://doi.org/10.1080/19345740802114533

Klein, A., Starkey, P., & Ramirez, A. B. (2002). *Pre-K Mathematics Curriculum*. Scott Foresman.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). *STEM: Good jobs now and for the future*. ESA Issue Brief# 03-11. US Department of Commerce.

Lange, J. de. (2007). Large-scale assessment and mathematics education. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1111–1142). Information Age Publishing.

La Paro, K. M., Hamre, B. K., LoCasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., Clifford, R., Barbarin, O., Howes, C., & Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education & Development*, *20*(4), 657–692. https://doi.org/10.1080/10409280802541965

Mashburn, A., Justice, L. M., McGinty, A., & Slocum, L. (2016). The impacts of a scalable intervention on the language and literacy development of rural pre-kindergartners. *Applied Developmental Science*, *20*(1), 61–78. https://doi.org/10.1080/10888691.2015.1051622

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., … Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x

Massachusetts Department of Education. (2003). *Guidelines for preschool learning experiences*. Massachusetts Department of Education. http://www.eec.state.ma.us/docs1/curriculum/20030401_preschool_early_learning_guidelines.pdf

McCormick, M. P., Cappella, E., O'Connor, E. E., & McClowry, S. G. (2015). Social-emotional learning and academic achievement: using causal methods to explore classroom-level mechanisms. *AERA Open*, *1*(3), 1–26. https://doi.org/10.1177/2332858415603959

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*(2), 93–127. https://doi.org/10.2307/1170709

Michigan State Board of Education. (2013). *Early childhood standards of quality for prekindergarten*. Michigan State Board of Education. https://www.michigan.gov/documents/mde/ECSQ_OK_Approved_422339_7.pdf

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*(1), 18–35. https://doi.org/10.3102/0013189X16633182

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making Pre-K Count: Improving math instruction in New York City* (October). Manpower Demonstration Research Corporation http://www.mdrc.org/sites/default/files/Making_Pre-K_Count_FR.pdf

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American journal of evaluation, 24*(3), 315-340. https://doi.org/10.1177/109821400302400303

Muthén, B. O., & Muthén, L. K. (2007). Mplus (Version 4.2) [Statistical software]. Muthén & Muthén.

National Association for the Education of Young Children. (1986). Position statement on developmentally appropriate practice in programs for 4- and 5-year-olds. In S.

Bredekamp (Ed.), *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (Expanded ed.), pp. 20–29. NAEYC.

National Association for the Education of Young Children (1996). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8: A position statement of the National Association for the Education of Young Children* (July). National Association for the Education of Young Children.

National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education (2009). Early learning standards: Creating the conditions for success. *Joint position statement*. www.naeyc.org/dap

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. National Governors Association.

National Research Council (1999) *Designing mathematics or science curriculum programs: A guide for using mathematics and science education standards.* National Academies Press, 2004. http://www.nap.edu/catalog/9658/designing-mathematics-or-science-curriculum-programs-a-guide-for-using

National Research Council (2012). *A Framework for K-12 Science Education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. https://doi.org/10.17226/13165

National Research Council (2013). *Next Generation Science Standards: For states, by states*. National Academies Press. http://www.nextgenscience.org/

National Research Council, Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials. (2004). *On evaluating curricular effectiveness: Judging the quality of k-12 mathematics evaluations*. National Academies. https://www.nap.edu/catalog/11025/on-evaluating-curricular-effectiveness-judging-the-quality-of-k-12

National Science Board. (2010). *Preparing the next generation of STEM Innovators: Identifying and developing our nation's human capital* (NSB Publication No. NSB-10-33). National Science Foundation. http://eric.ed.gov/?id=ED514061

National Science Board. (2015). *Revisiting the STEM workforce: A companion to science and engineering indicators 2014* (NSB Publication No. NSB-2015-10). National Science Foundation. http://www.nsf.gov/nsb/publications/2015/nsb201510.pdf

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services and Research*, *39*(4), 374–396. https://doi.org/10.1007/s11414-012-9295-x

Nilsson, P. (2008). Teaching for understanding: The complex nature of pedagogical content knowledge in pre-service education. *International Journal of Science Education*, *30*(10), 1281–1299. https://doi.org/10.1080/09500690802186993

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education and Development*, 24, 1000–1019. http://dx.doi.org/10.1080/10409289.2013. 825565

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in k-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793

Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., Horn, E., Palmer, S. and Marquis, J., (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly*, *25*(3), 314–328. https://doi.org/10.1016/j.ecresq.2010.03.001

Office of Head Start. (2010). *The head start child development and early learning framework: Promoting positive outcomes in early childhood programs*. U.S. Department of Health and Human Services. http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/eecd/Assessment/

Office of Head Start. (2015). Head Start Early Learning Outcomes Framework: Ages birth to five. U.S. Department of Health and Human Services. https://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf

Organisation for Economic Co-operation and Development (2016). *PISA 2015 Results (Volume 1)*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

Pence, K. L., Justice, L. M., & Wiggins, A. K. (2008). Preschool teachers' fidelity in implementing a comprehensive language-rich curriculum. *Language, Speech & Hearing Services in Schools*, *39*(3), 329–341. https://doi.org/10.1044/0161-1461(2008/031)

Penuel, W. R., Phillips, R. S., & Harris, C. J. (2014). Analysing teachers' curriculum implementation from integrity and actor-oriented perspectives. *Journal of Curriculum Studies*, *46*(6), 751–777. https://doi.org/10.1080/00220272.2014.921841

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0167660

Phillips, B. M., Ingrole, S. A., Burris, P. W., & Tabulda, G. (2017). Investigating predictors of fidelity of implementation for a preschool vocabulary and language curriculum. *Early Child Development and Care*, *187*(3–4), 542–553. https://doi.org/10.1080/03004430.2016.1251428

Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of childcare. *Early Childhood Research Quarterly*, *12*(3), 281–303. https://doi.org/10.1016/S0885-2006(97)90004-1

Pianta, R., Downer, J., Hamre, B., & Harare, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *The Future of Children*, *26*(2), 119–137. https://doi.org/10.1353/foc.2016.0015

Pianta, R. C., Nimetz, S. L., & Bennett, E. (1997). Mother-child relationships, teacher-child relationships, and school outcomes in preschool and kindergarten. *Early Childhood Research Quarterly*, *12*(3), 263–280. https://doi.org/10.1016/S0885-2006(97)90003-X

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Paul H. Brookes

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*(4), 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001

Piasta, S. B., Justice, L. M., McGinty, A., Mashburn, A., & Slocum, L. (2015). A comprehensive examination of preschool teachers' implementation fidelity when using a supplemental language and literacy curriculum. *Child & Youth Care Forum*, *44*(5), 731–755. https://doi.org/10.1007/s10566-015-9305-2

Piasta, S. B., Logan, J. A. R., Pelatti, C. Y., Capps, J. L., & Petrill, S. A. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms. *Journal of Educational Psychology*, *107*(2), 407–422. https://doi.org/10.1037/a0037621

Pinar, W. (2011). *The character of curriculum studies: Bildung, currere, and the recurring question of the subject*. Palgrave MacMillan.

Platz, Donald; Arellano, J. (2011). Time Tested Early Childhood Theories and Practices. *Education*, *132*(1), 54–63. http://proxy01.its.virginia.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=66538794&site=ehost-live&scope=site

Preschool Curriculum Evaluation Research Consortium (2008). *Effects of preschool curriculum programs on school readiness (NCER 2008-2009*). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Government Printing Office. http://ncer.ed.gov

Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., and Tang, J.H. (2016). *Highlights from TIMSS and TIMSS advanced 2015: Mathematics and science* *achievement of U.S. students in grades 4 and 8 and in advanced courses at the end of high school in an international context* (NCES 2017-002). U.S. Department of Education, National Center for Education Statistics. http://nces.ed.gov/pubsearch.

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2005). Optimal Design for longitudinal and multilevel research (Version 1.55) [Statistical software] University of Michigan.

Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry*, *29*(3), 315-342. https://www.jstor.org/stable/3185911

Rising Above the Gathering Storm Committee (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. National Academies Press. http://www.nap.edu/catalog.php?record_id=12999

Ruiz-Primo, M. A. (2006). A *multi-method and multi-source approach for studying fidelity of implementation*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2006). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *1*(2), 89–119. https://doi.org/10.1080/19345740801941332

Scott-Little, C., Lesko, J., Martella, J., & Miliburn, P. (2007). Early learning standards: Results from a national survey to document trends in state-level policies and practices. *Early Childhood Research and Practice*, *9*(1), 1–23. https://eric.ed.gov/?id=EJ1084857

Shepard, L. a., & Smith, M. L. (1988). Escalating academic demand in kindergarten: Counterproductive policies. *The Elementary School Journal*, *89*(2), 134. https://doi.org/10.1086/461568

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. http://www.jstor.org/stable/1175860

Sabol, T. J., Hong-Soliday, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can Rating Pre-K Programs Predict. *Science*, *341*, 845–846. https://doi.org/10.1126/science.1233517

Stipek, D. (2006). No Child Left Behind comes to preschool. *The Elementary School Journal*, *106*(5), 455–466. https://doi.org/10.1086/505440

Sutterby, J. A., & Kharod, D. (2020). Re-examining play in the early childhood curriculum. In J. J. Mueeler & N. File (Eds.), *Curriculum in early childhood education: Re-examined, reclaimed, renewed* (2nd ed., pp. 120-143). Routledge.

Superfine, A. C., Marshall, A., & Kelso, C. (2015). Fidelity of implementation: bringing written curriculum materials into the equation. *Curriculum Journal*, *26*(1), 164–191. https://doi.org/10.1080/09585176.2014.990910

Tao, Y., Oliver, M., & Venville, G. (2012). Long-term outcomes of early childhood science education: Insights from a cross-national comparative case study on conceptual understanding of science. *International Journal of Science and Mathematics Education*, *10*(6), 1269–1302. https://doi.org/10.1007/s10763-012-9335-2

Tu, T. (2006). Preschool science environment: What is available in a preschool classroom? *Early Childhood Education Journal*, *33*(4), 245–251. https://doi.org/10.1007/s10643-005-0049-8

U.S. House, Committee on Science (2001). *National Science Education Act of 2001* (H. Rpt. 107-133, Pt. 1). Government Printing Office. https://www.congress.gov/congressional-report/107th-congress/house-report/133/1

Vygotsky, L. (1997). Interaction between learning and development. In M. Gauvain & M. Cole (Eds), *Readings on the Development of Children*, (pp. 29-35). W. H. Freeman and Company. (Reprinted from Vygotsky, L. S. (1978)). *Mind in society: The development of higher psychological processes*. (pp. 79-91), Harvard University Press.

Wertsch, J. V. (1998). *Mind as action*. Oxford University Press.

Whittaker, J. V., Kinzie, M. B., Williford, A., & DeCoster, J. (2016). Effects of MyTeachingPartner–Math/Science on teacher–child interactions in prekindergarten classrooms. *Early Education and Development*, *27*(1), 110–127. https://doi.org/10.1080/10409289.2015.1047711

Xu, Y., Chin, C., Reed, E., & Hutchinson, C. (2014). The effects of a comprehensive early literacy project on preschoolers' language and literacy skills. *Early Childhood Education Journal*, *42*(5), 295–304. https://doi.org/10.1007/s10643-013-0613-6

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., … Zaslow, M. J. (2013). *Investing in Our Future: The Evidence Base on Preschool Education*, October. Society for Research in Child Development. http://fcd-us.org/resources/evidence-base-preschool

Zan, B., & Donegan-Ritter, M. (2014). Reflecting, coaching and mentoring to enhance teacher-child interactions in head start classrooms. *Early Childhood Education Journal*, *42*(2), 93–104. https://doi.org/10.1007/s10643-013-0592-7

Zaslow, M., Tout, K., & Martinez-Beck, I. (2011). *Measuring the quality of early care and education programs at the intersection of research, policy, and practice, OPRE research-to-policy, research-to-practice brief* (OPRE 2011-10a). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational research*, *86*(4), 981-1015.

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. I. Stamenov (Ed.), *Language structure, discourse, and the access to consciousness* (pp. 113-153). John Benjamin Publishing Company.

Zellman, G. L., Perlman, M., Le, V. N., & Setodji, C. M. (2010). *Assessing the validity of the Qualistar early learning quality rating and improvement system as a tool for improving child-care quality*. Rand Corporation. http://www.rand.org/pubs/research_briefs/RB9539.html

Zigler, E. (1987). Formal schooling for four-year-olds? No. *American Psychologist*, *42*(3), 254–260. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=ovfta&NEWS=N&AN=00000487-198703000-00007

APPENDICES

Appendix A

Fall Weekly Activities (W=Week, A=Activity)

September	October	November
W1A1-The Five Senses	W1A1-Body Parts I	W1A1-Animal Habitats I
W1A2-Senses:Sight	W1A2-Body Parts II	W1A2-Animal Habitats II
W2A1-Senses: Learning to Listen	W2A1-Soil I	W2A1-Living vs. Non-living I
W2A2-Senses: Feeling & Describing	W2A2-Soil II	W2A2-Living vs. Non-living II
W3A1-Senses: Smelling & Describing	W3A1-Worm I	W3A1-Human Food
W3A2-Senses: Tasting & Describing	W3A2-Worm II	W3A2-Animal Food
W4A1-Plants and Environments I	W4A1-Recycling & Reusing I	
W4A2-Plants and Environments II	W4A2-Recycling & Reusing II	

December
W1A1-Sky I
W1A2-Sky II: Sun
W2A1-Night Sky I
W2A2-Night Sky II
W3A1-Moon and Sun
W3A2-Daytime and Nighttime Animals

Spring Weekly Activities (W=Week, A=Activity)

January	February	March
W1A1-Motion I	W1A1-Floating and Sinking I	W1A1-Wind I
W1A2-Motion II	W1A2-Floating and Sinking II	W1A2-Wind II
W2A1-Solids and Liquids	W2A1-Magnets I	W2A1-Clouds I
W2A2-Water and Ice	W2A2-Magnets II	W2A2-Clouds II
W3A1-Changing Matter with Water I	W3A1-Shadows I	W3A1-Simple Tools I
W3A2-Changing Matter with Water II	W3A2-Shadows II	W3A2-Simple Tools II
W4A1-Animal Behavior: How We Stay Warm I	W4A1-Building Materials	W4A1-Seeds I
W4A2-Animal Behavior: How We Stay Warm II	W4A2-Building a Bridge	W4A2-Seeds II

April	May
W1A1-Stems, Roots, Leaves, and Seeds	W1A1-Plant Growth I
W1A2-Changing Habitats	W1A2-My Growth I
W2A1-Protecting Nature	W2A1-My Growth II
W2A2-Plants in Spring	W2A2-Plant Growth II
W3A1-Families I	W3A1-Insects
W3A2-Families II	W3A2-Born from Eggs
	W4A1-Observe Living Things
	W4A2-Metamorphosis

Appendix B

Sample MTP-Science Activity Guide Sheet

March-Science-W1-A1	**Wind I**	Whole Group

GET READY	**Objectives** • Describe what happens when the wind blows • Describe things that use wind	**Use the Lingo** • Move • Blow • Strong • Breeze • Sailboat • Kite • Windmill • Pinwheel
	Materials: • *The Wind Blew*, by Pat Hutchins • Pictures of things that need the wind to move or do not. **Preparation:** • N/A	
ENGAGE	1. Science Chant. 2. Simulate wind by using a book to move the air toward the students. If appropriate, ask the students to use paper or books to fan themselves and each other to move the air, which feels like wind. Ask them how they feel when the moving air blows on them. 3. *Let's read this book to find out more about wind!*	

INVESTIGATE

4. Read *The Wind Blew*.
   - On some of the pages, ask the students :
     - *How can you tell the wind is blowing?*
     - *What does the wind in the book feel like or sound like?*
     - *Does the wind in the book look like a mild breeze or strong wind?*
     - *Why are the items in the book up in the air? Why did they fall back to the ground at the end of the book?*
5. Sort pictures according to whether they need the wind to move or not. (For example, an ocean liner can be moved by the wind but it does not need the wind to move, however, a sailboat does.)
   - Ask students to explain the reasons for their sorting.
   - If students are unable to sort, prompt the students to think if wind can be used to move the depicted object or parts of it. Note: Boats with sails are designed to use the wind for movement, but all boats floating on the surface may be blown by the wind. Hair may be blown, but a person is not usually moved by the wind.

DISCUSS

6. Ask questions about the book and apply it to student experiences.
   - *Have you ever been outside on a windy day? What happened? What did it feel like?*
   - *Can you see the wind? How do you know it is there?*
   - *Do you like the wind? Always? Why? Why not?*

Use opportunities that arise throughout the day to talk about wind, what it feels like, and its effects. • Observe and record what they observe happening when the wind blows on a windy day. • Predict what stronger wind would do to houses, trees, and power lines, and help them find books with pictures that support, add to, and/or negate their predictions. • Dramatize the book with wind from a fan. • If the opportunity arises when wind is present during outside time, use the teachable moment to connect the wind (the way it feels and how things look blowing in the wind) to the book and the sorting pictures.	EXTEND

For Students With More Advanced Skills	**For Students Requiring More Support**	
• Refer to the book: Compare the weights of items that were blown away to those that were not. • During sorting, challenge students to find an object in the room that could be sorted along with the pictures.	• Allow students to use pictures in the book to assist them during sorting.	MAKE IT WORK

Appendix C

Fidelity Coding Manual – Year 3 – 1/26/10

myTeachingPartner Mathematics Science

Fidelity of Curriculum Implementation

Observation Checklist

Coder: ____________________ Start Time: ____________________

Tape Name/Number: ____________________ End time: ____________________

Activity Name: ____________________ Specified Activity Format: Whole Group Small Group Center

Instructions:

1. Read the teaching instructions for the given activity before watching the video clip.
2. Coding starts when activity starts, not when the tape starts.
3. Watch the entire activity, then enter a rating in response to the questions below.
4. When determining whether or not a code is "1," keep in mind that it is "all or nothing." The direction must be completed fully to get credit.

TEACHER'S FACILITATION

	NO	YES	NA	CONSTRUCT
1. Activity is conducted in specified format (whole group, small group).	0	1	NA	Activity Completion
Explanation: Activity setting is noted in the top right corner of the lesson plan. Whole group denotes all the children in the classroom (approximately 18 children per class). Small group is approximately half of the class (about 9 children). Even if they are at separate tables, if the TA is not implementing the lesson, then the activity is considered to be whole group because the students are only being instructed by the teacher and are not getting the same amount of personal attention.				
2. Teacher uses materials (or an equivalent substitution) as specified in the lesson plan.	0	1	NA	Materials & Environ
Explanation: Teacher uses materials provided by MTP or an equivalent or better substitution, in addition to those indicated as "teacher provided." An example of an equivalent substitution would be using a felt board in exchange for the laminated board provided by MTP. If materials are listed as "N/A" in the lesson, then indicate "NA" for this question. Even if you can see the materials in the background of the video, it only counts if the teacher actually uses them.				
3. Teacher begins lesson with the chant.	0	1	NA	Ensuring Engagement
Explanation: No explanation needed.				
4. Teacher introduces activity with the engage prompt posed in the curriculum.	0	1	NA	Content Coverage
Explanation: Teacher must use the prompt provided in the lesson plan to engage the children in order to score a "1". Note: The prompt is the numbered step and not necessarily the subsequent bullet point. Slight wording changes are acceptable but any changes to content will be scored a "0." Keep in mind that it is "all or nothing" coding.				

5. Teacher implements the "discuss" portion of the lesson plan at the appropriate time.	0	1	NA	Supporting Cognition & Lang
Explanation: The discuss portion of the lesson is implemented when it seems appropriate. This may or may not be at the end of the activity. If it is at the end of the activity, then it is fine but teachers may also use the discussion questions throughout the activity. If this is the case, the teacher will be coded as implementing the discuss section and will not be given credit for treating it as an adaptation.				

(Bold-faced word is the target for None/Some/Most/All)	None	Some	Most	All	NA	CONSTRUCT
Explanation: None= none of the children, none of the time, or not at all—0%. Some=1-50%. Most= 51-99%. All=all of the children all, all of the time, or completely—100%.						

6. Teacher actively models the use of math and/or scientific **language** (from "Use the Lingo").	1	2	3	4	NA	Supporting Cognition & Lang.
Explanation: The extent to which the teacher uses the words/phrases from "Use the Lingo" while implementing the lesson. Words used in book readings count in this section, as long as they are used in an appropriate context. For example, teacher's use of the word "more" will only count if it is used in reference to a specific math context, such as "amount."						
7. Children use the specified math and/or scientific **language** in their comments (from "Use the Lingo").	1	2	3	4	NA	Supporting Cognition & Lang.
Explanation: Children use the words from "Use the lingo" themselves. What percentage of words were spoken by 1 or more of the children?						
8. Teacher elicits children's observations and explanations for the activity, **during available opportunities**.	1	2	3	4	NA	Supporting Cognition & Lang.
Explanation: During available opportunities, the teacher asks children to make their own observations or explain concepts in their own words. For example, the teacher initiates "feedback loop" questioning with children to try to have the child extend upon his or her answers. Pay attention to missed opportunities and situations in which children raise their hands but are not called upon or try to share their opinion but are shushed.						

9. **Children** share their own observations and explanations	1	2	3	4	NA	Supporting Cognition & Lang.
Explanation: This is different from item 9 because this is just what the children are doing and does not take into account whether children's observations/explanation are encouraged by teacher. This refers to the number of children who speak at some point during the activity.						

10. Teacher extends/elaborates on students' **comments or questions.**	1	2	3	4	NA	Supporting Cognition & Lang.
Explanation: This is not necessarily teacher initiated (but can be). The teacher takes the students' comments or questions and uses them when appropriate to go back and review concepts, uses them to guide the discussion, adds vocabulary, repeats and extends student comments, etc.						
11. All **main components** (numbered steps) in the activity plan are completed.	1	2	3	4	NA	Content Coverage
Explanation: The extent to which the teacher implements each numbered step. The score is not lowered if a teacher skips a "For example" section. If the numbered step is not done in its entirety then it is counted as a skipped step.						
12. All **supporting components** (bullet points under numbered steps) in the activity plan are completed.	1	2	3	4	NA	Content Coverage
Explanation: The extent to which the teacher completes the bulleted points under each numbered step. If a bullet is not done in its entirety, then it is counted as a skipped step.						
13. Children perform the activity tasks specified in the lesson plan.	1	2	3	4	NA	Ensuring Engagement
Explanation: The extent to which the children participate in/perform the tasks described in the lesson. Do at least some of the children perform the tasks specified in the lesson plan. This question refers to the tasks in the lesson plan, whether or not the teacher asked the children to complete them. For example, if the teacher does not have the children participate in the way specified in the lesson plan, it will be scored down. This includes participation in discussion (for example. students answering questions from the lesson plan.) This may not necessarily be all the children completing each task but this is coded on the task level. To code this, step back from the activity and look at it globally to determine the number of tasks that children were asked to complete and the number they actually completed. Divide the number they completed by the total possible tasks for a percentage. It is important to take a high level view of the activity and think about what they children should be both doing and to what they should be offering responses.						
14. Teacher **explanations** of the mathematical or scientific concepts are accurate.	1	2	3	4	NA	Content Coverage
Explanation: Teacher explains scientific and mathematic concepts accurately. The purpose of this item is to detect those areas of common misconceptions in math and science as well and areas of content knowledge for which teachers are deficient. To score lower than a "4' the teacher must explain a concept inaccurately. For example, teacher discusses a common misconception as accurate. Teachers should be scored downward for allowing misconceptions to stand.						

Appendix D

This list of science assessment items was adapted from two types of original documents used by trained assessors—assessment response sheets and assessor script—so as to provide expedient context. Point values are noted in parentheses next to each item.

Life Science

Topic #1: Living vs. Non-Living Things

Part 1: Identifying Living vs. Non-Living Things

Shown a picture of a dog.

1. What is this? (1)
2. Are dogs alive or not alive? (1)

Shown a picture of a chair.

3. What is this? (1)
4. Are chairs alive or not alive? (1)

Given following pictures at once, sort living things into groups.

(Animate)

5. Blue Jay (1)
6. Butterfly (1)
7. Fish (1)

(Inanimate)

8. Pansies (1)
9. Tree (1)

Given following pictures at once, sort non-living things.

(Artifacts)

10. Bicycle (1)
11. Pencil (1)

(Natural Kinds)

12. Clouds (1)
13. Lightning (1)
14. Ocean (1)

Topic #2: Plants vs. Animals

Part 1: Animals

15. Does an animal need food to grow and be healthy? (1)
16. Does an animal need water to grow and be healthy? (1)
17. Does an animal need air to grow and be healthy? (1)
18. Does an animal need light to grow and be healthy? (1)

Part 2: Identifying Plants vs. Animals

Shown a picture of a tree.

19. What is this? (1)
20. Are trees plants or animals? (1)

Given 10 pictures, sort into plants and animals:

21. Plants (1)
22. Animals (1)

Topic #3: Parts and Functions of the Body

Part 1: The Senses

Shown a picture of a child's face, point to the body part that this child would use to

23. smell? (1)
24. see? (1)
25. hear? (1)
26. taste? (1)

Part 2: Parts of Animal Bodies

Shown a picture of a horse, then next a bird, identify body parts and point to parts associated with functions.

27. Horse's foot (1)
28. Horse's tail (1)
29. What horse uses to smell (1)
30. What horse uses to hear (1)
31. Bird's eye (1)
32. Bird's feet (1)

33. Bird's wings (1)

Shown a picture of a fish.

34. What fish uses to eat (1)
35. What fish uses to see (1)

Topic #4: Plant Biology

Part 1: Seeds

Shown pictures of fruits with seeds.

36. *Point to the seeds in each picture:* (1)
    - Watermelon
    - Apple
    - Cucumber
    - Strawberry

Part 2: Parts of Plants

Shown pictures of several plants and asked to point to…

37. Stem (1)
38. Roots (1)
39. Leaves (1)
40. Flowers (1)
41. Stem (1)

Shown a picture of one plant and asked

42. Does the plant need food to grow and be healthy? (1)
43. Does a plant need water to grow and be healthy? (1)
44. Does a plant need air to grow and be healthy? (1)
45. Does a plant need light to grow and be healthy? (1)

Part 3: Plant Life Cycles

Shown pictures of a green maple leaf and a red maple leaf and asked:

46. Which one of these leaves would you see outside in the Spring? (1)
47. Which one of these leaves would you see outside in the Fall? (1)

Topic #5: Animal Behavior During Day and Night

Shown a picture of an owl.

48. When do owls sleep, day or night? (1)
49. When do owls eat, day or night? (1)

Topic #6: Animal Families

Part 1: Baby-Adult Pairs

Shown five unsorted parent-baby animal pairs.

50. Match babies with their parents (1)
    - Dogs
    - Sheep
    - Gorillas
    - Lions
    - Ducks

Shown one pairing from above.

51. Point to the one that is : (1)
    - a baby
    - a parent

Earth & Physical Sciences

Topic #1: Scientific Tools

Part 1: Identification

Shown hand lens.

1. What is this? (1)

Part 2: Using Hand Lens

Shown hand lens, balance, and a picture of a tiny ladybug.

2. If I want to see something that is really small, which tool should I use? (1)

Given hand lens to use and a picture of a tiny ladybug.

3. What do you see in the picture? (1)

Part 3: Magnets

Shown magnets:

4. What are these? (1)

Part 4: Using Magnets

Shown three magnetic and three non-magnetic objects and asked which ones will stick to the magnet.

5. paperclip (1)
6. screws (1)
7. battery (1)
8. wooden block (1)
9. rubber ball (1)
10. plastic bear (1)

Topic #2: Weather

Part 1: Wind

Given a paper fan.

11. Use this to make wind. (1)

Part 2: Using Wind

Shown pictures of items and asked to point to things that need wind to move.

12. Windmill (1)
13. Computer (1)
14. Chair (1)
15. Wagon (1)
16. Kite (1)
17. Sailboat (1)

Topic #3: Temperature

Part 1: Seasonal Temperature

Shown a picture of two apartments.

18. Point to the picture that shows SUMMER. (1)
19. Point to the picture that shows WINTER. (1)

Part 2: Temperature Changes

Shown pictures of ice cubes.

20. What are these? (1)

Topic #4: Materials

Part 1: Material Composition

Shown six objects made of three different materials.

Point to which things are made out of wood.

21. Blue block	(1)
22. Door hanger	(1)
23. Spatula	(1)
24. School Bus	(1)
25. Fork	(1)
26. Strawberry	(1)

Point to which things are made out of plastic:

27. Blue block	(1)
28. Door hanger	(1)
29. Spatula	(1)
30. School Bus	(1)
31. Fork	(1)
32. Strawberry	(1)

Point to which things are made out of metal:

33. Blue block	(1)
34. Door hanger	(1)
35. Spatula	(1)
36. School Bus	(1)
37. Fork	(1)
38. Strawberry	(1)

Topic #5: Solids vs. Liquids

Part 1: Identification

Shown a block and bottle of water.

39. Point to one that is liquid.	(1)
40. Point to one that is solid.	(1)

Topic #6: Motion

Part 1: Predicting Speed

Shown a picture with two balls on inclines.

41. Which ball will be faster? (1)

Shown a different picture with two balls on inclines.

42. Which ball will be faster? (1)

Topic #7: Buoyancy

Part 1: Material

Given three buoyant and three non-buoyant objects and asked to show which ones will float if put in water.

43. Pencil (1)
44. Rubber ducky (1)
45. Styrofoam (1)
46. Rock (1)
47. Quarter (1)
48. Battery (1)

Topic #8: Day & Night

Part 1: Sun and Moon

49. When can you see the sun?
50. When can you see the moon?

Appendix E

Data Analysis Models for Each Research Question and Hypothesis

Research Question	Hypothesis	Analysis Model	Procedure
R1: Does fidelity of curriculum implementation (adherence, quality of delivery, and dosage) predict gains in children's science knowledge and skills over the preschool year?	R1H1: Adherence, dosage, and quality of delivery (*Emotional Support*, *Classroom Organization*, and *Instructional Support*) each have positive additive effect on children's gains in science learning outcomes	*Science Outcome Spring* $= \beta_0 + \beta_1$ *Science Outcome Fall* $+ \beta_2$ *Plus* $+ \beta_3$ *Pre-K Experience* $+ \beta_4$ *Teacher Education* $+ \beta_5$ *Maternal Education* $+ \beta_6$ *Family Income* $+ \boldsymbol{\beta_7}$ ***Adherence*** $+ \boldsymbol{\beta_8}$ ***Dosage*** $+ \boldsymbol{\beta_9}$ **Quality** $+ b_j + \varepsilon_{ij}$	Each of the following models examines the hypothesis, R1H1. Model 1 with *Emotional Support*: ***…β7 Adherence + β8 Dosage + β9 ES…*** Model 2 with *Classroom Organization*: ***…β7 Adherence + β8 Dosage + β9 CO…*** Model 3 with *Instructional Support* ***…β7 Adherence + β8 Dosage + β9 IS…***

Research Question	Hypothesis	Analysis Model	Procedure
R2: Does the association of one fidelity component with children's gains in science learning outcomes depend on levels of another fidelity component?	R2H1: The relation of adherence to gains in children's science learning outcomes is greater when teachers implement curriculum using higher quality of delivery (*ES*, *CO*, and *IS*). R2H2: The relation of dosage to gains in children's science learning outcomes is greater when teachers implement curriculum using higher quality of delivery (*ES*, *CO*, and *IS*). R2H3: The relation of adherence to gains in children's science learning outcomes is greater when teachers implement curriculum using higher dosage.	*Science Outcome Spring* = β_0 + β_1 *Science Outcome Fall* + ... + β_6 *Family Income* + $\boldsymbol{\beta_9}$ ***Quality*** + $\boldsymbol{\beta_{10}}$ ***Adherence x Quality*** + $\boldsymbol{\beta_{11}}$ ***Dosage x Quality*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage*** + b_j + ε_{ij}	Each of the following models examines three of the hypotheses (R2H1, R2H2, R2H3) at once, Model 1 with *Emotional Support* : ***…*** $\boldsymbol{\beta_9}$ ***ES*** + $\boldsymbol{\beta_{10}}$ ***Adherence x ES*** + $\boldsymbol{\beta_{11}}$ ***Dosage x ES*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage…*** Model 2 with *Classroom Organization*: ***…*** $\boldsymbol{\beta_9}$ ***CO*** + $\boldsymbol{\beta_{10}}$ ***Adherence x CO*** + $\boldsymbol{\beta_{11}}$ ***Dosage x CO*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage…*** Model 3 with *Instructional Support* ***…*** $\boldsymbol{\beta_9}$ ***IS*** + $\boldsymbol{\beta_{10}}$ ***Adherence x IS*** + $\boldsymbol{\beta_{11}}$ ***Dosage x IS*** + $\boldsymbol{\beta_{12}}$ ***Adherence x Dosage…***