Investigating the Feasibility of Preschool Interventions Targeting Children's Executive Functioning & Math Skills

by

Lillie Moffett

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education and Psychology)
in the University of Michigan
2021

Doctoral Committee:

Professor Frederick J. Morrison, Co-Chair
Associate Professor Christina Weiland, Co-Chair
Research Associate Professor Robin Jacob
Professor Priti Shah

Lillie Moffett

lillmoff@umich.edu

ORCID iD: 0000-0003-1233-5185

ACKNOWLEDGEMENTS

This dissertation study was funded in part by the Office for Planning Research and Evaluation Head Start Dissertation Grant (90YR0111). I was also personally funded by the National Science Foundation Graduate Research Fellowship Grant (1650114) and the Institute for Education Sciences Causal Inference Research in Education Policy Fellowship Grant (R305B150012).

In addition to my dissertation committee, I would like to thank Child 360 for providing me with a valuable internship opportunity to work with their partnering preschools and coaches, my peers in the Equity in Early Learning Lab for their support and feedback throughout graduate school, and, of course, the wonderful children and teachers I was able to work with every day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Given the well-documented relation between executive functioning (EF) and math skills in preschool, there is a surprising lack of evidence on early intervention approaches that have successfully, and consistently, impacted both EF and math skills. To confront this gap in the literature, Study 1 explored why an established EF intervention impacted children's math skills and not EF. Study II examined whether explicitly *combining* EF and math content may impact both children's math and EF skills. And finally, Study III examined the feasibility of combined math/EF content *in the classroom*, and identified barriers to coaching, teacher implementation, and teacher understanding of the intervention components.

More specifically, Study 1 performed a replication and extension analysis of the Tools of the Mind (Tools) program's lack of impact on children's EF skills but small impact on math skills (Head Start CARES study, Morris et al., 2014). This preschool classroom intervention—which emphasizes planning, scaffolded play, and peer turn-taking—is theorized to improve children's EF skills, yet a meta-analysis of this curriculum has revealed more consistent impacts on math than EF skills, and two other studies found that higher fidelity of implementation at the end of this intervention was associated with *fewer* student EF gains. I approached this perplexing finding by taking a closer look at the fidelity of implementation of specific components of the intervention to determine if the documented low levels of fidelity in the Head Start CARES study was consistent across intervention components and time of year, and if subsequent variability in this more fine-grained measurement may be explaining the impacts on math and

(lack thereof) on EF. Results illustrated that teachers who achieved high levels of fidelity on one component of the intervention (Play plans) within the first two months of implementation had students who made more gains in math. These teachers were also more likely to have a Bachelor's (BA) Degree. However, a subgroup analysis revealed that there was no impact of Tools on math when the sample was restricted to teachers with and without a BA degree, although teachers with a BA in the Tools group had children who made more (trending towards statistically significant) gains in math compared to children who had teachers with a BA in the control group ($\beta = .12$, $p = 0.09$). This study implies that the documented impact of Tools on math skills is still not well-understood, although there is some evidence that high fidelity of implementation early on is contributing to the observed differences in children's math gains.

Study II was more exploratory than Study I where I designed an explicitly combined math and EF preschool intervention. Although results from Study 1 indicated that interventions targeting EF skills may not be a promising approach for impacting math *and* EF skills, it is possible that there is a way to design the content of EF interventions to impact both EF and math simultaneously. In this study, I created a set of 12 activities that explicitly targeted children's math and EF skills. Pilot data revealed that 40 children who participated in a 7-week intervention where they played EF/math games with a researcher outside of the classroom made statistically significantly greater gains in EF skills (d=.15)—but not math skills (d=.07)—compared to children who were not taken out of the classroom to play these games.

In Study III, I brought activities from Study II into the classroom setting to identify the strengths and weaknesses of the content and coaching of the intervention from the perspective of preschool teachers and coaches. I trained ten Head Start preschool teachers and four instructional coaches on the implementation of these activities during large- and small-group settings.

Although I collected no child-level measures, I used these feasibility data in preparation for a randomized-controlled trial (that is beyond the scope and funding of this dissertation). Observational data revealed that after 16 weeks of modeling, implementation, and feedback, teachers were able to implement these activities in the classroom with adequate fidelity, and children were rated as exhibiting mid to high levels of engagement on 10 out of 12 activities. When observed, teachers struggled the most with achieving fidelity in following the math skill sequence of each of the activities, despite implementing the EF integration with ease. However, when teachers were tested on their understanding of math skill sequence and EF integration for each of the activities, teachers struggled the most with identifying how children's EF skills were actually exercised within each activity. This has implications for how we educate teachers on the components of an intervention. Teachers' fidelity of implementation may also not reflect what they actually *understand* about the intervention. Qualitative data further highlighted strengths of the intervention and coaching where teachers lauded the in-classroom modeling of activities, breakdown of math skill levels for each activity, and variety of math topics covered. Noteworthy suggestions from teachers and coaches for future iterations of these math and EF activities included more coaching on how to help children with the most basic math skill level, aligning activities with the Desired Results Developmental Profile (DRDP) assessment and teachers' lesson plans, and aligning coaching with Classroom Assessment Scoring System (CLASS) goals.

Overall, this dissertation highlights the complexity of how early interventions impact EF and math skills and discusses the implications for research and practice.

CHAPTER I: Unpacking Existing Math and EF Preschool Interventions

Early childhood is characterized as a critical period for building foundational cognitive skills that have implications for how children positively contribute to future society (Karoly, 2002; Nelson, 2000; Shonkoff & Phillips, 2000). As such, increased policy attention has been given to developing early childhood programs and interventions to support children's learning. Despite these efforts, however, the field is still striving to fully understand how to intervene and impact children's cognitive skills, how to effectively coach teachers, and which approaches and skills—given limitation of time and resources—should receive the most attention. In the preschool literature, specifically, there are several evidence-based interventions, curricula, and programs targeting a range of children's skills from language and literacy (Justice, Mashburn, Hamre, & Pianta, 2008; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011; Schickedanz & Dickinson, 2005), to math (Presser, Clements, Ginsburg, & Ertle, 2015; Starkey, Klein, & Wakeley, 2004; Sarama & Clements, 2004), and socio-emotional skills (Domitrovich, Cortes, & Greenberg, 2007). Yet there is still fundamental uncertainty on *why* and *when* these interventions work, and *how* to best support teachers and encourage educators to replace less efficacious curricula with ones that are more beneficial for children.

In this dissertation, I narrowed my focus to preschool interventions that target children's executive function (EF) and math skills. I highlighted these two skills because other early interventions of children's skills—such as literacy and social-emotional skills—have been widely established and are more easily understood by preschool teachers in both their importance and theory of change (Lee & Ginsburg, 2007; Stipek, 2013). For EF skills, there is more mixed evidence on the overall impact of EF-targeted preschool interventions (Jacob & Parkinson, 2015) and minimal research on teachers' understanding of EF (in elementary school; Bardack & Obradovic, 2019). For math, there is more resistance from preschool teachers to teach this subject (Copley, 2004), and less understanding by teachers on the development of this skill in young children (Sarama, DiBiase, & Clements, 2004; Lee, 2017). As such, improving efforts on the design and implementation of preschool EF and math interventions is of timely importance in the field.

Executive Functioning and Math in Early Childhood

Defining executive functioning and early math learning. Researchers typically refer to executive functioning (EF) as comprising processes involved in intentionally controlling ones' impulses, attention, thinking, and behavior (Best, Miller, & Naglieri, 2011; Huzinga, Dolan, & van der Molen, 2006; Raver, 2013). EF consists of three primary components—working memory, cognitive flexibility, and response inhibition—that have all independently predicted gains in academic achievement (Nguyen & Duncan, 2019). Working memory is defined as the ability to hold and manipulate verbal or non-verbal information in one's mind (Baddeley, 1992; Welsh, 2002). Cognitive flexibility refers to the ability to alternate attention or response strategies as a task changes (Zelazo, Muller, Frye, & Marcovitch, 2003). And response inhibition involves suppressing prepotent responses that may otherwise be inappropriate for the task one is doing (Barkley, 1997).

All three of these components are crucial during the preschool years as they co-develop with children's academic learning and help children function in a classroom environment. For example, children's working memory helps children keep directions and rules in mind, keep track of and comprehend what they are reading, and hold quantities in their mind when learning to add and subtract. Children's cognitive flexibility can help them successfully transition and switch from one subject area to another and listen for nuances in teacher directions that might be different from the normal schedule. Response inhibition can help a child focus on a task while shutting out environmental distraction, inhibit the impulse to blurt out rather than raise their hand, and control the desire to play with instructional materials. Although separate EF components are highly correlated in preschool, there is some evidence for distinct working memory and inhibition skill (Lerner & Lonigan,2014), and unique prediction of cognitive flexibility, inhibition and working memory for children's math and literacy skills in preschool (Purpura, Schmitt, & Ganley, 2017). As such, in this dissertation, I discuss EF in terms of the separate components that comprise this construct.

When it comes to "defining" preschool math, the most widely emphasized math skills in early childhood consist of number sense knowledge such as numeral recognition, verbal counting, one-to-one correspondence, cardinality, counting-on, comparing magnitudes, and adding and subtracting quantities of one or two (Baroody, 2009). However, preschool-aged children are also developmentally equipped to learn math topics related to patterning, geometry, measurement, and aspects of algebraic thinking (Clements & Sarama, 2004), although these topics are less often taught by preschool teachers.

Early math learning typically involves the use of manipulatives to aid children in developing an understanding of quantity and concrete observation of adding and subtracting. Children also rely on manipulatives to improve their understanding of number, shapes, and patterns. In the preschool classroom, this can look like children counting bears and then matching this quantity to the written numeral, or matching a shape to a numeral indicating the number of sides of the shape. More unstructured early math learning can include using pretend money during dramatic play (e.g., buying pretend ice cream for “2” dollars), or asking children how many more cups are needed at the table during lunch time. Similar to EF skills, there is an established literature on the development of math in early childhood, and the pioneers on this topic—Drs. Douglas Clements and Julie Sarama—have conducted several research projects to enhance understanding of children’s math development and have since summarized their findings in their “Learning Trajectories Approach” to early math learning (Clements & Sarama, 2004).

Overall, EF and math development constitutes a relatively “new” topic on preschool teachers’ radars, and the two are often introduced to teachers separately in detailed learning standards or goals for children’s development. For example, in the California Preschool Learning Foundations (2008), executive functioning is couched in social-emotional learning and behavior regulation—which although not theoretically incorrect—immediately creates a disconnect for teachers from how this skill may be nurtured to support children’s math learning. If EF and math interventions and coaching are to be improved, providing teachers with a strong knowledge of the relation between the two skills is essential.

Theoretical and empirical evidence: Relation between EF and math. Research on EF in the context of schooling has perhaps generated the most robust empirical evidence for EF’s strong relation with math achievement (Blair & Razza, 2007; Blair et al., 2011; Blair, Ursache, Greenberg, Vernon-Feagans, & The Family Life Project Investigators, 2015; Fuhs, Farran & Nesbitt, 2015; McClelland et al., 2014). The connection between EF and math is intuitively appealing because even early math places demands on children’s working memory, cognitive flexibility, and response inhibition (Welsh, Nix, Blair, Bierman, & Nelson, 2010).

For example, empirical evidence has highlighted the importance of cognitive flexibility for children's patterning ability (Bock et al., 2015), as children need to be able to flexibly "switch" between different patterning units (e.g., AB vs. AAB). However, working memory and response inhibition have perhaps demonstrated the strongest relation with math achievement. Children practice response inhibition when they have to suppress the first solution that comes to mind or ignore salient extraneous information in a math word problem. One study with three-to-five-year-old children found that performance on a magnitude comparison task related to math achievement, but this relation was predominantly driven by trials that required inhibiting an irrelevant stimulus to select the larger number (Fuhs and McNeil, 2013). Furthermore, inhibitory control has also demonstrated associations with magnitude comparisons and overall math achievement in three to six year olds (Merkley, Thompson & Scerif, 2016). Working memory is also used during math tasks, since a majority of math involves maintaining, manipulating, and adding relevant information while simultaneously engaging in another cognitive demanding task. For example, adding and subtract requires children to keep quantities in their mind (e.g., mental math; "You have five pencils…now add two… now take away three" etc.). Past work has also shown that first graders' visual-spatial working memory predicted their ability to make magnitude judgments and correctly add and subtract (Simmons, Willis, & Adams, 2012). Performance in inhibition and working memory tasks in elementary school children has even illustrated predictions to math achievement, over and above earlier math abilities (Bos, Van der Ven, Kroesbergen, & Van Luit, 2013). There is also further evidence to suggest that a lack of inhibition and working memory leads to difficulty shifting and evaluating new strategies for dealing with math tasks in an elementary school sample (Bull & Scerif, 2001).

There is also research supporting the developmental sensitivity of relying on EF for success on math tasks. For example, one study illustrated that when the "EF load" varies in a magnitude comparison task, three year olds are worse at the task when their EF was taxed the most, whereas four year olds performed well across the "EF loads" (Prager, Sera & Carlson, 2016). This suggests that preschool children naturally recruit EF skills when engaging in math tasks. However, since this study was conducted with middle to high SES children, it is not known if low SES children also implicitly recruit EF during math tasks, or if they instead need support in explicitly strengthening EF skills to help solve a math problem. Thus, it is important to explore this phenomenon further with lower SES children (e.g., Head Start populations).

Despite the emergence of theoretical and empirical bidirectional relations between math and EF, it is unclear which skill strengthens the other. There is evidence from longitudinal studies that early EF contributes to later academic learning but not vice versa (Best et al., 2011; Bull et al., 2008, 2011; Clark et al., 2010, 2013; LeFevre et al., 2013; Steele et al., 2012). Conversely, other studies have illustrated that longitudinally, EF predicted math *and* math predicted EF (McClelland et al., 2007; Van der Ven, Kroesbergen, Boom, & Leseman, 2012; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Thus, the field is still unclear with how EF and math interact in young children to support both skills, so dissecting why some EF and math interventions impact these skills and some do not can be crucial to understanding this interplay and the further design of interventions.

Interventions Designed to Support EF and Math

Despite the documented relations between EF and math skills in early childhood, interventions of these skills have typically been approached separately. In this section, I use the term "EF-only" to characterize the interventions that have primarily focused on strengthening EF skills, and use the term “Math-only” to represent evidence-based curricula that are intended to impact children’s math skills.

EF-only interventions. Perhaps the most well-known preschool EF intervention is The Tools of the Mind (Tools) comprehensive curriculum, which is based on the Vygotskyan approach aimed at supporting children to be “masters of their own behavior” (Vygotsky, 1962, p. 147). Tools’ theory of change is comprised of the teacher regulating the students, the students regulating each other, and then the students self-regulating (Bodrova & Leong, 2007). A major component of this intervention is encouraging make-believe play, where children must negotiate roles, focus on their specific role, and inhibit the impulse to switch roles. Teachers encourage children to “plan out” their play in order to support goal-oriented thinking. Despite the explicit focus on encouraging children to recruit EF skills during play, evaluations of the Tools program in preschool have demonstrated small or null effects on executive functioning and academic achievement outcomes (*d*’s<.10; Barnett et al., 2008; Farran & Wilson, 2013; Morris et al., 2014), although there is some “promise” of this intervention in supporting children’s math skills.

The Red Light, Purple Light (RLPL) intervention—although smaller in scale than Tools of the Mind—also explicitly exercises children’s EF skills (Tominey & McClelland, 2011). This 8-week intervention contains 6 circle-time games that children participate in twice a week for 30 minutes. These games explicitly target children’s working memory, response inhibition, and cognitive flexibility. For example, in the “Freeze Game,” children are instructed to dance slowly to fast songs and quickly to slow songs, and then perform the opposite action where they dance slowly to slow songs and quickly to fast songs. An initial evaluation of this intervention demonstrated no treatment effect on EF, except for children of the lowest EF scores (effect size= .34). However, a RLPL evaluation in a lower-income sample (Head Start) than the original evaluation revealed an intervention effect on EF for the whole sample (*d*=.32; Schmitt et al., 2015). Additionally, when RLPL was combined with a kindergarten readiness intervention (Bridge to Kindergarten) for low-income children with no preschool experience, children in the intervention group made statistically significant more gains in EF compared to Bridge to Kindergarten classrooms that did not receive the RLPL component (*d*=.34; Duncan et al., 2018). This suggests that for children of lower-income backgrounds, the RLPL intervention can be helpful in promoting EF skills.

Other interventions supporting children's EF and related skills have focused more on general teacher practices instead of developing an explicit curriculum or activities targeting children's EF. For example, the Head Start REDI program fostered teachers' classroom management techniques and other general strategies to promote emotion and behavior regulation. This program also involved a curriculum (Promoting Alternative THinking Strategies (PATHS)) focused on language skills, literacy skills, emotional understanding, and self-regulation. However, evaluations of this PATHS program in preschool have illustrated small, positive impacts on children's EF skills (Bierman et al., 2008; Domitrovich, Cortes, & Greenberg, 2007; Morris et al., 2014), and literacy skills (Bierman et al., 2008; Morris et al., 2014). Another intervention—the Chicago School Readiness Program (CSRP)—provided teacher professional development on classroom management techniques that were theoretically aimed at providing teachers with the tools to effectively regulate children's behavior in the classroom (in addition to a mental health consultation component) (Raver et al., 2008). An evaluation of the CSRP in preschool demonstrated statistically significant impacts on children's response inhibition (*d*= .37), language (*d*=.34), literacy (*d*=.63) and math skills (*d*=.54; Raver et al., 2011). The main content for the teacher trainings in the CSRP intervention was based on the Incredible Years (IY)

Child Training curriculum, which was originally developed to support children with severe conduct problems and diagnosed oppositional defiant disorders (Webster-Stratton & Hammond, 1997; Webster-Stratton & Reid, 2003). In addition to the statistically significant impact of CSRP on children's EF, only one other evaluation of the teacher IY program in a non-clinical population has examined child EF outcomes and found that the IY intervention improved children's engagement in learning tasks and teachers' use of behavior management techniques, but this did not transfer to gains in children's EF skills (d= -.04; Morris et al., 2014).

For EF-only interventions in preschool, the overarching consensus is that they have small or null impacts on children's academic achievement (Jacob & Parkinson, 2015), and for impacts on EF skills, there is some evidence that they benefit children with low baseline EF skills, dual language learners, and children of lower SES (Tominey & McClelland, 2011; Raver et al., 2011; Schmitt et al., 2015). In order to determine whether or not the field should continue to pursue EF-only interventions, it is important to isolate whether the problem is that the theory of change (e.g., exercising/scaffolding EF will impact EF) is simply not working, or if the fidelity of implementation tools used are faulty in some sense. For example, most EF-only interventions have reported moderate-to-high levels of fidelity of implementation, which at a glance, creates the assumption that the former hypothesis is correct e.g., EF-only approaches do not work. However, fidelity of implementation measures used in EF-only interventions typically focus on dosage and adherence (e.g., Red Light, Purple Light and Tools of the Mind), but do not capture any nuance with implementation of content such as teacher implementation quality, children's ability to participate with the "EF" content, or whether or not teachers' fidelity of implementation is congruent with their understanding of the content and theory of change.

In the context of education interactions, many studies have highlighted the distinction between teacher knowledge and teacher practice in several domains such as literacy (McCutchen et al., 2002), math (Ball & Bass, 2002), and science (Goodnough, 2010), as well as the importance of teacher knowledge in predicting intervention success (Hamre et al., 2012). But no EF intervention study to date has provided qualitative data on teacher knowledge or understanding of EF content. For example, in the Red Light, Purple Light intervention, although authors reported general ease for teachers in implementing a "stop/go" game, they did not report on whether or how the teachers were educated on the theory of change of this intervention. They also did not report whether teachers understood how to move children through a developmental continuum of this skill. This understanding of child development within a specific domain is a crucial component of teacher knowledge (Ben-Peretz, 2011), and is used extensively in the design and coaching of the Building Blocks preschool math curriculum (Clements & Sarama, 2007). In an analysis of the Tools of the Mind curriculum, authors do attempt to collect data on teacher interest in *gaining knowledge* of the curriculum's purpose by examining how often a teacher asked about the program and how the program helps children. However, findings illustrated that teachers rarely asked these kinds of questions after the initial training (only 14% of teachers; Morris et al., 2014). This lack of data on teacher knowledge or understanding could be a potential hole in EF-intervention evaluations, and something that needs to be explored before concluding that exercising EF is not an efficient means of strengthening EF.

Math-only interventions. In contrast to EF-only interventions, there is much more promising research on math-only interventions. Currently, there are four evidence-based math curricula in preschool settings. The most widely known is Building Blocks (Clements & Sarama, 2007, 2008), which has demonstrated statistically significant impacts on children's growth in math skills (*d* = .72 on REMA; Clements et al., 2011; Clements et al., 2013) and EF skills (*d*=.21; Clements et al., 2012). This curriculum has even demonstrated effectiveness at large scale when combined with an evidenced based literacy curriculum (OWL) (*d*'s = .20- .27 for EF; *d*'s= .49 - .58 for math; Weiland et al., 2013). Building Blocks has a strong coaching model and curriculum framework that spans multiple contexts of instruction (e.g., whole-class, small-group, computer stations), and math skills (e.g., geometry, numeracy, patterning, measurement, data analysis). On average, Building Blocks requires 21-26 minutes of math instruction in the school day (Clements & Sarama, 2008; Clements et al., 2011; Sarama et al., 2008). A unique feature of Building Blocks is that it contains an individualized component that helps teachers recognize and move children along a specific developmental trajectory. Additionally, when Building Blocks was supplemented with a follow-up intervention in kindergarten (High 5's Math Club), children who received this supplement and the preschool Building Blocks (compared to children who only received Building Blocks in preschool, but not this supplement) made statistically significant gains in math (*d*=.19), but not EF (*d*'s=-.03-.05). And children who participated in both interventions made statistically significantly more math gains (*d*=.30) than children who participated in neither intervention (again, no impact on children's EF skills; *d*'s = .08) (Jacob, Erickson & Mattera, 2018). This illustrates that Building Blocks is an efficacious option for strengthening children's math (and in some cases, EF) skills, but continued math support following this intervention exposure is important for sustaining intervention effects.

Other evidenced-based math curricula include Big Math for Little Kids (BMLK; Ginsburg, Greenes, & Balfanz, 2003; RCT by Presser, Clements, Ginsburg & Ertle, 2015), The Pre-K Mathematics Curriculum (PMC; Klein et al., 2008), and Numbers Plus (Wakabayashi et al., 2020). All of these curricula have also demonstrated statistically significant impacts on child math outcomes (BMLK effect size = .32 on ECLS-B assessment; PMC effect size =.55 on Child Mathematics Assessment; Numbers Plus effect size = .14 on EMAS). These curricula also follow a proposed scope and sequence that align with developmental learning theories in math development as well as Head Start Child Development and Early Learning Framework guidelines. Similar to Building Blocks, these curricula cover topics spanning number concepts & quantities, number relationship & operations, geometry & spatial sense, patterns, and measurement & comparison. They also offer small-group instruction within each math domain and BMLK offers an additional whole-class component. However, evaluations of these curricula included control groups that were doing very little math instruction to begin with, so it is unclear if impacts are due simply to increased dosage or the quality and specific facets of the curricula themselves. The fact that each of these evidenced-based math curricula use different child math assessments—and have never been systematically compared to each other—creates difficulty in determining which curricula works best. Additionally, these three math curricula have not been empirically evaluated for their impact on EF skills.

EF and math interventions. As presented thus far, there is promising evidence for the effectiveness of math interventions in impacting math and EF skills, and less compelling evidence for the utility of EF interventions impacting these skills. However, considering the intertwined nature of EF and math, it is possible that combining both of these skills in an intervention may support each other.

Clements and colleagues (2016) argued that EF may be developed through the learning of math, and “a more intentional development of math curricula based on recent research on EF may do both even more effectively” (p. 86). There is some evidence for the utility of this combined approach in elementary school with computer trainings that integrate academic content with attention training vs. attention training alone (Capraro, Capraro, & Rupley, 2011; Iseman & Naglieri, 2011; Rabiner et al., 2010). However, in the preschool literature there is only nascent evidence on the impacts of this approach, and results are mixed.

Evaluations that have compared math interventions to combined math-EF interventions highlight how focusing on math alone can improve EF skills, whereas programs that have tried to combine the two have not been particularly successful. For example, in one evaluation, Building Blocks (math-only) showed statistically significant impacts on EF and math, compared to a Building Blocks + Tools Curriculum (d=.21 EF, d=.23 math; Clements, Sarama, Unlu & Layzer, 2021). In a separate evaluation with a low-income sample, Building Blocks only outperformed the Building Blocks + Tools curriculum on children's EF skills, but not math skills (Clements et al., 2020; d=.19). And in a follow-up intervention of The Red Light, Purple Light (RLPL) where authors added math and literacy content in the music and movement "stop/go" games (McClelland et al., 2019), both the RLPL-only and the RLPL + math/literacy condition impacted children's math skills compared to a business as usual group (d=.31, d=.38) (but not literacy or EF skills).

In summary, there is no compelling evidence for the effectiveness of combining math and EF content in the context of preschool interventions. However, to date, most studies have used existing interventions to combine EF and math, but only a few have developed an intervention from the ground up. For example, Connect4Learning is a preschool curriculum designed to integrate cognitive and social-emotional processes, such as executive functioning, with literacy and STEM related subject materials. Similarly, in 2018, the Chan Zuckerberg Initiative proposed A Bold Initiative to Development Breakthrough Solutions in Executive Function and Math, which included a call for “creating curricula with a focus on formative feedback and explicitly integrating social-emotional learning with academic content” (pg. 12) (Gates Foundation, 2019).

This dissertation contributes to this effort in identifying key ingredients of EF content that may be supporting children’s math skills (Study 1), pilot evidence for a combined EF and math content approach with preschoolers outside of the classroom (Study II), and the feasibility of this approach with teachers and coaches in the classroom (Study III).

References

Arnold, D.H., Fisher, P.H., Doctoroff, G.L., & Dobbs, J. (2002). Accelerating math development in Head Start classrooms. *Journal of Educational Psychology, 94*(4), 762-770.

Baddeley, A. (1992). Working Memory. *Science, 255*(5044).

Bardack, S., & Obradovic, J. (2019). Observing teachers' displays and scaffolding of executive functioning in the classroom context. *Journal of Applied Developmental Psychology, 62*, 205-219.

Baroody, A.J. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development, 1*, 80-128.

Barkley, R.A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94.

Barnett, W.S., Jung, K., Yarosz, D.J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effets of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*(3), 299-313.

Best, J.R., Miller, P.H., & Naglieri, J.A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327-336.

Blair. C., & Razza, R.P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child*

Development, 78(2), 647-663.

Blair. C., Protzko, J., & Ursache, A. (2011). Self-regulation and the development of early literacy. In D. Dickinson & S Neuman (Eds.), *Handbook of early literacy research* (pp. 20-35). New York: Guilford Press.

Blair, C., Ursache, A., Greenberg, M., Vernon-Feagans, L., & The Family Life Project Investigators. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter-word knowledge in the early elementary grades. *Developmental Psychology, 51*(4), 459-472.

Bock, A., Cartwrigth, K.B., Gonzalez, C., O'Brien, S., Robinson, M.F., Schmerold, K., Shriver, A., & Pasnak, R. (2015). The role of cognitive flexibility in pattern understanding. *Journal of Education and Human Development, 4*(1), 19-25.

Bodrova, E., & Leong, D.J. (2007). Play and Early Literacy: A Vygotskian Approach. From Play in Literacy in Early Childhood: Research From Multiple Perspectives.

Boonen, A.J.H., Kolkman, M.E., & Kroesbergen, E.H. (2011). The relation between teachers' math talk and acquisition of number sense within kindergarten classrooms. *Journal of School Psychology, 49*(3), 281-299.

Bos, I. F., Van der Ven, S.H.G., Kroesbergen, E.H., & Van Luit, J.H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29-44.

Bull, R., Epsy, K.A., & Wiebe, S.A. (2008). Short-term memory, working memory and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205-228.

Bull, R., Epsy, K.W., Wiebe, S.A., Sheffield, T.D., & Nelson, J.M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679-692.

Capraro, R.M., Capraro, M.M., & Rupley, W.H. (2011). Reading-enhanced problem solving: A theoretical model. *European Journal of Psychology of Education*, *27*(1), 91-114.

Clark, C.A.C., Pritchard, V.E., & Woodward, L.J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176-1191.

Clark, C.A.C., Sheffield, T.D., Wiebe, S.A., & Epsy, K.A. (2012). Longitudinal Associations Between Executive Control and Developing Mathematical Competence in Preschool Boys and Girls. *Child Development*, *84*(2), 662-677.

Clements, D.H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks Project. *Journal for Research in Mathematics Education, 38*(2), 136-163.

Clements, D. H., Sarama, J., Unlu, F., & Layzer, C. (2012, March). *The efficacy of an intervention synthesizing scaffolding designed to promote self-regulation with an early mathematics curriculum: Effects on executive function.* Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Clements, D.H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36*, 79-90.

Clements, D.H., & Sarama, J. (2008). Experimental Evaluation of the Effects of a Research-Based Mathematics Curriculum. *Educational Research Journal, 45*(2) 443-494.

Clements, D.H., Sarama, J., Wolfe, C.B., & Spitler, M.E. (2013). Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies: Persistence of Effects in the Third Year. *American Educational Research Journal, 50*(4). 812-850.

Clements, D.H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36*(3) 79-90.

Clements, D.H., Sarama, J., Layzer, C., Unlu, F., & Fesler, L. (2020). Effects on Mathematics and Executive Function of a Mathematics and Play Intervention Versus Mathematics Alone. *Journal for Research in Mathematics Education*, *51*(3) 301-333.

Copley, J. V. (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and kindergarten mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Domitrovich, C.E., Cortes, R.C., & Greenberg, M.T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28*(2), 67-91.

Farran, D. C., Lipsey, M. W., & Wilson, S. J. (2011, November). *In Curriculum and pedagogy: Effective math instruction and curricula*. Paper presented at the early childhood math conference, Berkeley, CA.

Farran, D.C., Wilson, S.J., & Lipsey, M. (2013). Effects of a curricular attempt to improve self-regulation and achievement in prekindergarten children. *Society for Research on*

Educational Effectiveness.

Fuhs, M., & McNeil, N.M. (2013). ANS acuity and mathematics ability in preschoolers from low income homes: contributions of inhibitory control. *Developmental Science, 16*(1), 136-148.

Gates Foundation (2019). EF + Math Program. *NewSchools Venture Fund*. Oakland, CA.

Ginsburg, H.P., Greenes, C.E., & Balfanz, R. (2003). *Big math for little kids: Pre-kindergarten*. Pearson Learning Group.

Huzinga, M., Dolan, C.V., & van der Molen, M.W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11) 2017 – 2036.

Iseman, J.S., & Naglieri, J.A. (2011). A cognitive strategy instruction to improve math calculation for children with ADHD and LD: a randomized controlled study. *Journal of Learning Disabilities, 44*(2), 184-195.

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*(4), 512-552.

Justice, L.M., Mashburn, A.J., Hamre, B.K., & Pianta, R.C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*(1) 51-68.

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

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a Pre-Kindergarten Mathematics Intervention: A Randomized Experiment. *Research on Educational Effectiveness, 1*(3), 155-178.

Lee, J.E. (2017). Preschool Teachers' Pedagogical Content Knowledge in Mathematics. *International Journal of Early Childhood, 49*, 229-243.

Lefevre, J. A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, *4*, 1-9.

Lonigan, C.J., Farver, J.M., Phillips, B.M., & Clancy-Menchetti, J. (2011). Promoting the development of emergent literacy skills: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24,* 305-337.

Mattera, S., Jacob, R., & Morris, P. (2018). *Strengthening Children's Math Skills with Enhanced Instruction: The Impacts of Making Pre-K Count and High 5s on Kindergarten Outcomes.* New York: MDRC.

McClelland, M.M., Cameron, C.E., McDonald, C., Farris, C.L., Jewkes… Morrison, F.J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947-959.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Frontiers in Psychology, 5*, 1-14.

McClelland, M.M., & Cameron, C.E. (2019). Developing together: The role of executive function and motor skills in children's early academic lives. *Early Childhood Research Quarterly*, *46*(1) 142-151.

Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology, 7*.

Nguyen, T., & Duncan, G.J. (2019). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly, 46*(1), 49-61.

Prager, E.O., Sera, M.D., & Carlson, S.M. (2016). Executive function and magnitude skills in preschool children. *Journal of Experimental Child Psychology, 147,* 126-139.

Presser, A.L., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big Math for Little Kids: The Effectiveness of Preschool and Kindergarten Mathematics Curriculum. *Early Educational and Development, 26*(3) 399-426.

Rabiner, D.L., Murray, D.W., Skinner, A.T., & Malone, P.S. (201). A Randomized Trial of Two Promising Computer-Based Interventions for Students with Attention Difficulties. *Journal of Abnormal Child Psychology, 38*, 131-142.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K., & Sardin, L. (2008). Improving pre-school classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10–26.

Raver, C.C., McCoy, D. C., Lowenstein, A. E., & Pess, R. (2013). Predicting individual differences in low-income children's executive control from early to middle childhood. *Developmental Science*, *16*(3), 394–408.

Raver, C.C., Jones, S.M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's Impact on Low-Income Preschoolers' Preacademic Skills: Self-Regulation as a Mediating Mechanism. *Child Development, 82*(1), 362-378.

Rudd, L. C., Lambert, M. C., Satterwhite, M., & Smith, C. H. (2009). Professional development + coaching = enhanced teaching: Increasing usage of math mediated language in preschool classrooms. *Early Childhood Education Journal*, *37*(1), 63–69.

Sarama, J., DiBiase, A., & Clements, D. (2004). *The professional development challenge in preschool mathematics* in Engaging Young Children in Mathematics.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Sarama, J., & Clements, D.H. (2013*). Lessons learned in the implementation of the TRIAD scale-up model: Teaching early mathematics with trajectories and technologies*. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.) *Applying implementation science in early childhood programs and systems* (p.173 – 191). Paul H Brookes Publishing Co.

Schacter, J., Shih, J., Allen, C.M., DeVaul, L., Adkins, A.B., Ito, T., & Jo, B. (2015). Math Shelf: A randomized trial of a prekindergarten tablet number sense curriculum. *Early Education and Development, 27*(1), 74-88.

Schickedanz, J., & Dickinson, D. (2005). *Opening the world of learning*. Iowa City, IA: Pearson Education

Schmitt, S.A., McClelland, M.M., Tominey, S.L., & Acock, A.C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly, 30*, 20-31.

Simmons, F.R., Willis, C., & Adams, A. (2012). Difference components of working memory have different relationships with difference mathematical skills. *Journal of Experimental Child Psychology, 111*(2), 139-155.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*(1), 99-120.

Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: Differential developmental gateways to literacy and numeracy. *Child Development, 83*(6), 2028-2041.

Tominey, S.L., & McClelland, M.M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education and Development, 22(*3), 489-519.

Van der Ven, S.H.G., Kroesbergen, E.H., Boom, J., & Leseman, P.P. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*(2), 100-119.

Vygotsky, L.S. (1962). Thought and language. Cambridge, MA. MIT Press.

Wakabayashi, T., Andrade-Adaniya, F., Schweinhart, L.J., Xiang, Z., Marshall, B.A., & Markley, C.A. (2020). The impact of a supplementary preschool mathematics curriculum on children's early mathematics learning. *Early Childhood Research Quarterly, 53*(4), 329-342.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65*(1), 93-109.

Webster-Stratton, C., & Reid, M. J. (2003). The incredible years parents, teachers and children training series: A multifaceted treatment approach for young children with conduct problems. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for* *children and adolescents* (pp. 224-240). New York, NY, US: The Guilford Press

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

Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In D. L. Molfese & V. J. Molfese (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139-185). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Welsh, J.A., Nix, R.L., Blair, C., Bierman, K.L., & Nelson, K.E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43-53.

Zelazo, P.D., Muller, U., Frye, D., Marcovitch, S… Carlson, S.M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development, 68*(3), 151-157.

CHAPTER II: Why Does the Tools of the Mind Intervention Impact Math Skills but not EF Skills?

This chapter explores an evaluation of the Tools of the Mind (Tools) curriculum to determine key active ingredients of the intervention and differences between teachers that may have been influencing impacts on children's math and EF gains. I use the existing Head Start CARES dataset to study this phenomenon given the diversity of the teacher sample in this dataset, as well as the rich, monthly fidelity of implementation data available. The original Head Start CARES analysis did not examine how teachers or differences in fidelity of implementation accounted for impacts on EF and math, so this chapter sheds light on this unexplored area.

Theory of Change

The Tools of the Mind (Tools) curriculum was developed over 25 years ago and has since been utilized in 18 states and 160 school districts, as well as in Canada (Bodrova & Leong, 2007; Bodrova & Leong, 2018). The hallmark of Tools' activities is the emphasis on children's development of self-regulation and executive functioning—which the creators of Tools highlight as the foundation of higher mental functions (Vygotsky, 1987). In Tools classrooms, children are taxed with exercising their self-regulation as they engage in different roles during make-believe play, take turns during peer reading, and stick to a plan for their play during center time. The creators of Tools relied heavily on Vygotskian theory in the design of this curriculum and in generating the specific "tools" children need to develop. The first "tool" is self-talk or "private speech," which helps children regulate their behaviors (Berk, 1992, Luria, 1969, Vygotsky, 1987). Another tool is encouraging the use of external mediators that help children carry out an intended action. For example, in adulthood, we rely on lists to lessen our mental load and maps to get us from point A to B. In Tools classrooms, children create "play plans" that help them remember their goal during center time and hold pictures of "lips read, ears listen." reminders are until they internalize the rules of peer reading. Due to the theory of change of the Tools of the Mind curriculum, it has been consistently referred to as a "self-regulation" and "EF" intervention in the literature.

Empirical Evidence

To date, there have been some positive evaluations of the Tools of the Mind curriculum in preschool and kindergarten. Most of these studies have found positive impacts on teachers' overall classroom practices such as improvement in scaffolding (part of Tools; *d*=.37, Morris et al., 2014) and classroom management techniques (*d*=.38, Barnett et al., 2008), as well as a decrease in teacher-reported student problem behaviors (*d*=-.50, Barnett et al., 2008; *d*=-.19, Blair, McKinnon, & Daneri, 2018; *d*=-.43, Diamond, Lee, Senften, Lam, & Abbot, 2019). However, only a few studies have also demonstrated statistically significant impacts on children's EF and academic achievement skills, although positive findings to date are only observed in sub-samples of low-income students (EF *d*=.12, math *d*=.13, literacy *d*=.14, Blair & Raver, 2014; EF *d*=.28, *d*=.31, Diamond et al., 2007), racial-ethnic minorities (literacy *d*=.22, Millaway, 2015), and dual-language learners (language *d*=.25, Cohen, Kramer-Vida, Frye, & Andreou, 2014; language *d*=.29, Daneri, Blair, Hammer, & Lopez, 2017).

In contrast, there is far more evidence for the *lack of efficacy* of the Tools curriculum on children's EF and academic achievement outcomes. For example, Clements, Sarama, Unlu and

Layzer (2012) found that a mathematics curriculum had stronger impacts on children's EF skills compared to Tools of the Mind combined with this math curriculum (part of Tools; *d*=.34). Lonigan and Phillips (2012) found that a similar approach of combining literacy activities with Tools of the Mind produced stronger impacts on children's literacy skills than Tools of the Mind alone (*d*=.24) or a business as usual group (*d*=.17), although no statistically significant impacts on EF were observed for any condition (*d*=-.02). Another study in preschool also found that Tools did not impact EF (*d*=.04; Solomon et al., 2017) compared to another play-based intervention. And perhaps one of the most comprehensive evaluations of Tool of the Mind to date illustrated that the curricula in preK did not impact children's gains in math, literacy or EF, and in fact led to less self-regulation growth in kindergarten (*d*=-.22; Farran & Wilson; 2012; 2014) compared to a business as usual group. Baron et al., (2020) also reported less gains in self-regulation in a sample of preK, kindergarten and 1st grade students (*d*=-.14, *p*=.08) compared to a business-as-usual group. For impacts on math, a meta-analysis of six datasets, including a few of the aforementioned studies, concluded that the Tools curriculum in preschool statistically significantly improved math skills, but effect sizes were small (*d*=.05). Effect sizes for Tools' impact on EF (*d*=.12) and literacy (*d*=.02) were, on average, non-significant (Baron et al., 2017). In summary, the literature to date is mixed on Tools' impact on children's skills in preschool and kindergarten, but there is some evidence suggesting that there may be a small impact on math skills that has yet to be fully understood.

The "Black Box" of the Tools of the Mind Curriculum

As previously emphasized, components of the Tools' intervention should not theoretically directly impact children's math skills. Yet, empirically, there are some positive impacts on children's math skills. One hypothesis is that an EF-focused intervention may influence children's math skills due to the intertwined nature of EF and math (Schmitt et al., 2017), but as previously highlighted, there are few studies that have actually found impacts on EF in the first place. So, the indirect impact on children's math skills from the impact on children's EF skills is not a likely explanation. A second hypothesis is that there is something about the specific components of Tools of the Mind that is driving this impact on math skills. For example, although Tools of the Mind primarily emphasizes "tools" for pretend play, the intervention includes other facets that may be improving children's math skills. To date, few studies have examined the impact of fidelity to specific components (beyond pretend play) of the intervention on child outcomes. A third hypothesis is that there is something about how the teachers are responding to the intervention that may in turn influence their math instruction. For example, perhaps the strategies provided to the teachers in the curriculum—such as scaffolding children's ideas—permeate teachers' math instructional practices. Unfortunately, evaluations of the Tools of the Mind curriculum have not typically collected data on these carryover effects of an intervention either via observation or teacher report. Yet, one way to address this hypothesis is to investigate whether there is something unique about the teachers themselves that may have driven the impact on math skills. For example, more educated or experienced teachers may be more likely to internalize the theory of change of the intervention and readily apply it to other areas of their teaching. To date, no Tools' evaluation has looked at this subgroup analysis and determined whether differences in teachers could have accounted for the observed impact on math skills. Given the diversity in teacher characteristics and rich fidelity of implementation data in the Head Start Cares evaluation of the Tools of the Mind curriculum, I pursue these latter two hypotheses in depth.

The fidelity of implementation hypothesis. I first consider fidelity of implementation as a factor in explaining intervention impacts. Fidelity to the intervention components refers to how well the intended intervention was carried out by the teacher and responded to by the students and consists of four primary components. Adherence is how well the teacher adheres to the intervention content as detailed in a given training manual. Dosage is how long students received an intervention and if they received the intended sequence. Quality refers to how well the curriculum is delivered. And participant responsiveness is how well the students respond to the intervention as observed by their participation, enthusiasm, and attentiveness (Gresham 2009; Power et al., 2005).

Theoretically, researchers expect teachers who achieve higher levels of fidelity to the intervention components to also have larger impacts on their children (Hulleman & Cordray, 2009), which has been supported with empirical evidence (Mendive, Weiland, Yoshikawa, & Snow, 2015; Humphrey, Barlow, & Lendrum, 2018). However, there is mixed evidence as to whether fidelity of implementation to the intervention core components—or "active ingredients"—accounts for observed intervention impacts on children's language and literacy skills in preschool specifically (Piasta et al., 2015; Reinke et al., 2013) and which facets of fidelity matter most in explaining this variance (McGinty et al., 2011). For example, teachers who achieved high levels of fidelity to one component of the intervention did not necessarily score highly on another (Hamre et al., 2010; Pianta et al., 2008) so this difference in fidelity scores could have accounted for different child impacts observed between classrooms. Furthermore, differences in fidelity of implementation of specific components within Tools specifically could also aid in deciphering the documented mixed fidelity of implementation (e.g., some studies reporting overall high levels of fidelity, some overall low; Bodrova & Leong, 2019;

Barron et al., 2020; Farran & Wilson, 2014; Morris et al., 2014), which could be due to variation at the curricular component level.

In the Tools of the Mind curriculum, fidelity of implementation is typically characterized by adherence and quality. Namely, the adherence to certain components of the intervention (e.g., play plans, pretend play) and the teachers' quality (e.g., degree of scaffolding) of these components. However, the Tools of the Mind's theory of change is unique in that the nature of high-quality implementation should theoretically change throughout the intervention as teachers' scaffolding decreases when children begin to "master [their] own behavior" and self-regulate (Vygotsky, 1987). As a result, the fidelity of implementation tool used in the beginning of the intervention should be different than the measure used at the end of the intervention, and this might also differ by fidelity to different components in the intervention that take longer to adopt (Bickman et al., 2009). Yet the indicators of fidelity used in most evaluations of Tools remain the same throughout the intervention period, and authors have not typically examined how fidelity to different parts of the intervention matter. This lack of sensitivity in the fidelity tool across time might explain how one study noted that teachers who had higher fidelity of implementation at the end of the intervention had children who made statistically significantly fewer gains in self-regulation (nearly a quarter of a standard deviation) compared to teachers who had lower levels of fidelity of implementation (Farran & Wilson, 2014). Similarly, Baron et al. (2020) found that teachers achieving high levels of fidelity of implementation on Make-believe play at the end of the year also led to fewer child gains in self-regulation (d=-.12). Neither study observed this phenomenon for math outcomes, although the association between fidelity and math gains was also negative.

The Head Start CARES data set is unique to study this first hypothesis because it provides *monthly* fidelity of implementation on *specific* components of the Tools' curriculum. Most evaluations have only collected three time points, reported average scores across components, or only focused on Make-believe play. In the Head Start CARES dataset there is also substantial variability in the fidelity of implementation to different components across time. As evidenced later, around a third of teachers steadily increased in fidelity of implementation over time, a third steadily decreased in fidelity of implementation over time, and a third stayed the same (e.g., low, medium, or high) in their fidelity of implementation over time. Thus, this variation in fidelity scores across time may have explained variation in children's outcomes. I hypothesize that since the fidelity measure itself did not change throughout the course of the intervention, teachers who achieved high levels of fidelity earlier on in the intervention (compared to teachers who achieved high levels of fidelity later) would predict gains in children's EF and math skills.

The teacher characteristics hypothesis. Another factor that might have explained the impact of Tools on math skills was due to differences between teachers, specifically how different teachers have interpreted or implemented the intervention. For example, prior work has theorized that less experienced teachers are also typically better implementers of interventions in general because they have higher motivation for career advancement and are more open and flexible to changing their teaching (Hildebrandt & Eom, 2011; Klimes-Dougan et al., 2009). Similarly, more educated teachers are theorized to be better implementers because they have a higher level of prior knowledge that assists with their understanding and application of the professional development they receive (Landry, Anthony, Swank, & Monseque-Bailey, 2009). In the preschool literature specifically, more educated preschool teachers (e.g., those with a

Bachelor's degree or higher) typically have larger impacts on children (Landry et al., 2006). However, there is contrasting evidence for neither teacher education nor years of teaching experience contributing to the fidelity of implementation of a literacy curriculum (Wasik & Hindman, 2011).

To date, no study has examined child impact differences or differences by fidelity of implementation by teacher characteristics in the context of the Tools intervention. Most evaluations of Tools have notreported years of teaching experience of the Tools teachers, and most studies have reported that all teachers have had at least a Bachelor's degree (Blair, McKinnon, & Daneri, 2018; Diamond et al., 2019; Farran & Wilson, 2014). Specifically, Diamond and colleagues (2019) emphasized how "Tools may work best with teachers with at least a Bachelor's education" (pg. 20). Other Tools evaluations have not mentioned teachers' education at all (Barnett et al., 2008; Baron et al., 2020; Clements et al., 2012; Diamond et al., 2007; Solomon et al., 2018), which makes it difficult to understand the mixed impacts between these studies and if they were due to differences between teachers.

In the current study, the Head Start CARES dataset is unique in that only 56% of teachers in both the Tools and Control group had a Bachelor's degree or higher, permitting examination of treatment impacts by teacher education. Similar to the fidelity data offered by the Head Start CARES data, this substantial variation in teacher characteristics may have also explained variation in children's outcomes.

Current Study

In the present study, I examined 1) the fidelity of implementation hypothesis e.g., differences in teachers' fidelity of implementation is accounting for the impacts on EF and math and 2) the teacher characteristic hypothesis e.g., differences in teachers' education and years of experiences is accounting for the impacts on EF and math, by exploring the Head Start CARES study (Morris et al., 2014). The purpose of the Head Start CARES study was to compare the impact of three different social-emotional preschool enhancements on 1) teachers' practices, 2) the climate of the classroom, 3) children's behavior regulation, executive functioning skills, emotion knowledge, and social problem-solving skills, and 4) children's preacademic skills. The three chosen enhancements were the Preschool PATHS (Promoting Alternative Thinking Strategies) curriculum, the Incredible Years Teaching Training Program, and Tools of the Mind-Play Curriculum.

The theory of change for each of the enhancements was that the coaching, teacher training, and classroom implementation would impact changes in teacher practices which would improve classroom quality and children's social-emotional and EF skills. These impacts may then have also indirectly supported children's math, literacy, and language skills. For Tools of the Mind specifically, Morris and colleagues hypothesized that coaching and implementation would impact teachers' improvement in scaffolding practices and children's literacy gains, improvement in emotion knowledge and social problem-solving, reduced problem behaviors, and gains in executive functioning.

Results revealed that teachers in Tools of the Mind classrooms (compared to control classrooms) did engage in more scaffolding practices (d=.68) and children had increased emotion knowledge (d=.12) at the end of preschool. However, there were no observed impacts on executive functioning (d=.00), reduced problem behaviors (d=.02) or social problem-solving skills (d=.04). There was a trending towards statistically significant impact of Tools of the Mind on children's math skills (d=.09) and no impacts on literacy (d=.00). Morris and colleagues postulated that their math finding may have been due to teachers' increased use of more small-group activities in Tools classroom, which may best support math learning.

There were three major differences in impact findings between the Tools intervention and the other two enhancements (Incredible Years and PATHS). One, only Tools had an impact on math. Two, Tools did not have an impact on EF (Incredible Years did). And three, Tools achieved only "satisfactory" levels of fidelity of implementation, compared to high fidelity reported by Incredible Years and PATHS. These major differences support the hypothesis that there might be key elements of fidelity of implementation that are impacting children's math skills (and not EF skills). Examining teacher characteristics will also help untangle whether any fidelity of implementation differences are due to differences between the teachers themselves.

Research Questions

1) What is the impact of Tools of the Mind on children's EF and math skills? (Replication)

2) Does average fidelity, initial fidelity, and end fidelity across the intervention components (Tools of the Mind group only) predict children's gains in EF and math skills?

3) Do teacher education and years of teaching experience predict teacher implementation? Are there differences in impacts on children's math and EF skills on by teacher education and experience?

Method

Random Assignment

As paraphrased from the original Head Start CARES study (Morris et al., 2014), a total of 17 Head Start grantees located in 10 states were selected to reflect the geographic, racial, and ethnic diversity of the national Head Start population. Cohort One consisted of 4 grantees in the Northeast that participated during the 2009-2010 school year, and Cohort Two consisted of 13 grantees from the rest of the country that participated in the 2010-2011 year. Random assignment was conducted within groups of similar centers within grantees. Centers within each block (within grantees) were then randomly assigned to one of four groups: Incredible Years, Preschool PATHS, Tools of the Mind-Play, or the Control group. The Control group was a "business as usual" group where no coaching was offered beyond what the classroom or center may have already been receiving. There were a total of 104 centers across 22 blocks: 18 blocks included 4 centers in the study and 4 blocks included e8 centers. Within each center, there were up to six classrooms and an average of nine children per classroom included in the study.

Eligibility & Sample Demographics

Children were eligible for the study if they were four years old in the fall. English or Spanish was their primary language, and they were not foster children. At baseline, 79% of eligible children had parents who consented to let them participate in the study. At follow-up, an average of 90% of eligible children had parents who consented. In the current study, I restricted the sample to the Tools of the Mind intervention group and the Control group. For this specific sample, there were a total of 26 centers, 76 classrooms (and teachers), and 678 children in the Tools of the Mind group and 26 centers, 77 classrooms (and teachers), and 621 children in the Control group who participated. Out of a total 52 coaches in the Head Start CARES study, 19 coaches participated in the coaching of Tools of the Mind teachers.

Since the original study looked only at baseline differences between all three interventions and the control group, I analyzed baseline differences between the Tools of the Mind group and Control group, specifically. Baseline differences were calculated using the What

Works Clearinghouse Procedure where mean differences between Treatment and Control group were divided by the pooled standard deviation of the construct (WWC Version 4.1 Procedures Handbook, 2020). There were no statistically significant differences between Tools of the Mind group and Control group on child age (effect size=.03), baseline math scores (effect size=.07), baseline Head-to-Toes (effect size=.01), or baseline Pencil Tap (effect size=.01). For teacher characteristics, there were no statistically significant differences between Tools and Control group on percentage of teachers who had a Bachelor's degree or higher (effect size=.02). However, there was a statistically significant difference between teachers in the Tools and Control group on years teaching preschool (effect size=.11)—with more teachers in the Tools group having more than 10 years of teaching experience. Results and an overall F-test of baseline characteristics are presented in Appendix A Tables 1-2.

Descriptive statistics on the demographics of the sample were calculated but crosschecked with the original study. On average, across the Tools and Control groups, 48% of the sample was female, children's mean age was 4.41, 19% of children were White, 26% were Black, 48% were Hispanic, and 9% identified as "other/multiracial." For the teacher characteristics of interest, 57% of teachers had a Bachelor's degree or higher in the Tools group and the Comparison group, and 74% and 68% of teachers in Tools and Control group had at least 10 years of teaching experience, respectively. For coach demographics, 100% of coaches were female, with a mean age of 46 years, 92% white (7% Hispanic, 1% other), and 100% had a master's degree in early childhood education and a minimum of 11 years of experience in early childhood settings (*M*= 22.30 years).

Procedure

The following procedure details were paraphrased from the original Head Start CARES study (Morris et al., 2014). For both intervention and control groups, baseline information on teachers (teacher self-survey) was collected in the spring prior to implementation (e.g., before summer training and introduction to Tools of the Mind). In the fall, informed consent forms for children were collected. Direct assessment data, teacher reports on children, and parent survey data were collected after consents were obtained (September to December). Follow-up data were collected for children in the spring of the implementation year (March to May).

Coaches (a mix of part-time and full-time employees of the grantee) were trained for one day before the start of the intervention year by trainers and developers for the respective enhancements. Coaches were supervised by trainers with monthly check-ins throughout the intervention year and there was on average one coach per 6 classrooms. On average, coaches visited each classroom 1x per week and spent anywhere between 1 hour to 2.5 hours in each classroom (M=1.2 hours). During these visits, coaches filled out weekly logs to document the amount of time they spent in observation or modeling the classroom (M=46 minutes) and what was covered in their one-on-one weekly meeting with the teachers (M=22 minutes). Monthly, coaches rated fidelity of implementation in Tools classrooms which was aggregated to create an overall average for the year.

Teachers were expected to attend enhancement trainings throughout the school year and their respective coach attended the trainings with them. Teachers attended two days of training in August/September, one day in October, one day in January, and one day in March (five days total). A day of training lasted between six and eight hours, and new content was presented at four out of the five training sessions. Overall, lead teacher attendance was relatively high across the trainings for the Tools group (92%).

Measures

Child-level

***Executive function*.** There were two measures of children's EF skills. The Head-to-Toes task assessed children's working memory and inhibitory control skills. In this task, children were instructed to touch their head when the assessor directed them to touch their toes, and then touch their toes when the assessor directed them to touch their head. Children were scored on the number of trials they answered correctly out of 10 trials. This activity was intended to tap into children's suppression of a dominant response in order to carry out a subdominant response. This task has demonstrated reliability and validity (with teacher ratings of self-control and academic skills; McClelland et al., 2007; Ponitz et al., 2008). In the second EF measure, the Pencil Tap task, children were asked to tap on a table twice with a pencil when the assessor tapped once, and once when the assessor tapped twice. Children were scored on the proportion of trials answered correctly out of 16 trials. Similar to the Head-to-Toes task, this task also required children to inhibit a natural tendency to mimic the action of the assessor while remembering the rule for the correct response. This task has also demonstrated reliability and validity with Head Start populations and has been shown to be predictive of math ability (Blair & Razza, 2007; Lipsey et al., 2017; Raver et al., 2011).

***Pre-academic skills*.** The Woodcock Johnson Applied Problems III subtest (W-score) was used in the fall and spring to measure children's math skills (Woodcock, McGrew, & Mather, 2001). This assessment had children identify numbers and quantities, and has demonstrated strong validity (McGrew & Woodcock, 2001). The Expressive One-Word Picture Vocabulary Test (EOWPVT) was also used in the fall and spring to assess children's expressive vocabulary in English and Spanish (Brownell, 2000). This task had children produce the word that best described the picture. Both the original study and I used the standard score of this assessment as a covariate in all child-level analyses.

Classroom-level

Teacher self-report survey. Teachers reported on their education and years of teaching experience (preschool-aged children). Binary variables (0/1) were created for each teacher characteristic. For teacher education, a 0 indicated that a teacher did not have a Bachelor's degree and a 1 indicated that a teacher had a Bachelor's or higher. For teaching experience, a 0 indicated that a teacher had less than 10 years of teaching preschool children and a 1 indicated that a teacher had 10 or more years of teaching experience. Although a 5-year split is typically used in the literature (Kraft, 2014), 90% of teachers had more than 5 years of teaching experience, so 10-year split provided more variation (~68% of teachers had more than 10 years of teaching experience).

Fidelity of implementation. Coaches used coaching weekly (and monthly) fidelity logs to document their impressions of teachers' implementation of the core components of the Tools of the Mind intervention. Items were scored on a scale of 1 (low) to 5 (high). Fidelity was measured in terms of both quality and dosage (frequency and duration) and was rated on a scale of 1 (low) to 5 (high) and achieving a rating of 3 or above was considered to be "above the threshold for implementing satisfactorily." The core components of fidelity of implementation are listed in Table II.1. To create an "initial" fidelity score, September and October scores were averaged. To create an "end" fidelity score, March and April scores were averaged (I created these "initial" and "end" variables). And average fidelity was an average of fidelity scores, monthly, from September to April (original study created this variable).

Analytic Approach

Missing data. There were complete data on child-level baseline measures for 74% of the Tools sample and 71% of the Control group. All baseline data (Pencil Tap, Head-to-Toes, Woodcock Johnson Applied Problems) were imputed using multiple imputation in SAS (data were imputed by the original Head Start CARES team). There were complete data for initial fidelity of implementation because if teachers were missing a September or October fidelity score (22 out of 76 teachers were missing one or the other), instead of averaging the Sep/Oct score, I used only their non-missing fidelity score. This decision was appropriate because October and September fidelity scores were generally correlated with one another (r=.72). There were also complete data for average fidelity of implementation because the original study reported calculating average fidelity with just the non-missing points. There was a 96% response rate for the teacher survey for teachers in the Tools group and 93% response rate for teachers in the Control group. Due to the low rate of missingness for this measure, I used complete case analysis for models including teacher characteristics as a predictor.

Attrition. There were similar rates of attrition (child left sample after baseline data collection) in the Tools group as in the Control group (15%). There was also no differential attrition between the Tools and the Control group as there were no statistically significant differences between stayers and leavers within each condition and between conditions. This overall attrition and differential attrition rate was considered “low attrition” by IES standards (U.S Department of Education, 2003). The original study reported child demographic differences in attrition for all three enhancements and the Control group, so I conducted attrition analyses for children only in the Tools enhancement and the Control group. For statistically significant differences, there were slightly more boys who stayed in the sample (e.g., more girls left the sample) (effect size= -.14). Children who left the sample also had slightly lower Pencil Tap scores (effect size= -.14), and Woodcock Johnson Applied Problems scores (effect size=-.10). However these differences were only trending towards significant (p<.10). These differences by attrition are similar to the original study's attrition analysis for all three enhancements and the Control group, except that differences in Pencil Tap scores in the larger attrition analysis reached conventional levels of statistical significance. Findings from attrition analyses are in Appendix A Table 3.

Intracluster correlations. For the Pencil Tap outcome, 0% of the variance in children's scores was attributed to differences between classrooms and 1% was attributed to differences between centers. For the Head-to-Toes outcome, 5% of the variance in children's scores was attributed to differences between classrooms and 4% could be attributed to differences between centers. For the Woodcock Johnson Applied Problems outcome, 12% of the variance in children's scores was attributed to differences between classrooms and 24% was attributed to differences between centers. For fidelity of implementation, 27% of the variance in teachers' scores was attributed to differences between centers. Due to the variance at the classroom and center level, for models at the child level I used classroom and center random intercepts (and a fixed effect for block for impact analyses), and for models at the classroom-level I used center random intercepts.

Multi-level modeling. For Research Question 1 (Equation 1), I conducted a multi-level analysis with random intercepts for center and classroom and a fixed effect for block which replicates the analytic approach of the original study.

Equation (1)

$$Y_{skcb} = \beta_0 + \beta_1 T_{cb} + \beta_2 Z_b + \sigma X_{skcb} + e_{skcb} + u_{kcb} + v_{cb}$$

Where subscript *s* is student, subscript *k* is classroom, subscript *c* is center, and subscript *b* is block (center).

Y_{skcb} = the outcome of interest (Pencil Tap raw score, Head to Toes raw score, Woodcock Johnson Applied Problems W score)

T_{cb} = the treatment indicator, which equals 1 if center *c* was randomized to Tools of the Mind in block *b*, and 0 if it was randomized to Control

Z_b = an indicator variable for random assignment block *b*

X_{skcb} = a vector of baseline characteristics (child baseline measure of outcome, child age, and child baseline expressive language)

e_{skcb} = a random error for student *s* from classroom *k* in center *c* that is independently and identically distributed across students

u_{kcb}= a random error for classroom *k*

v_{cb}= a random error for center *c*

For Research Question 2 (Equation 2), I fit multi-level models and restricted the sample to Tools of the Mind classrooms (Control classrooms did not have fidelity of implementation data) and regressed child EF and math scores on fidelity of implementation at different times points, to different components, and included a random intercept for center.

Equation (2)

$$Y_{skc} = \beta_0 + \beta_1 Fidelity_{skc} + \sigma X_{skc} + e_{skc} + u_{kc}$$

Y_{skc} = the outcome of interest (Pencil Tap raw score, Head to Toes raw score, Woodcock Johnson Applied Problems W score)

$Fidelity_{skc}$ = the score on each Tools component that teachers received in 1) September/October "Initial," 2) March/April "End," and 3) "Average" (1-5)

X_{skc} = a vector of baseline characteristics (child baseline measure of outcome, child age, and child baseline expressive language)

e_{skc} = a random error for student *s* from classroom *k* in center *c* that is independently and identically distributed across students

u_{kc}= a random error for classroom *k*

For Research Question 3a, in equation 3 I restricted the sample to teachers without a Bachelor's degree to predict teachers' fidelity of implementation, and then re-fit models for teachers with a Bachelor's degree or higher (I did the same for years of teaching experience). And for Research Question 3b (subgroup analysis), I used the same model as Research Question 1 (Equation 1) but restricted the sample by teacher education (and did the same for years of teaching experience), and then created a combined model where teacher education (and teaching experience) interacted with the treatment indicator.

Equation (3)

$$Fidelity_{kc} = \beta_0 + \beta_1 Teach_{kc} + u_{kc}$$

$Fidelity_{skc}$ = the score on each Tools component that teachers received in 1) September/October "Initial," 2) March/April "End," and 3) "Average" (1-5) for teacher (classroom) *k* in center *c*

$Teach_{kc}$ = a 1/0 indicator for whether a teacher had a Bachelor's degree or higher, or no Bachelor's degree, for teachers in center *c* (Same for teaching experience > 10 years, 1/0 indicator)

u_{kc}= a random error for classroom *k*

Results

The descriptive statistics presented in this section replicated those presented in the Head Start CARES analyses. However, the original report did not present data on initial or end fidelity of implementation ratings so I could not crosscheck descriptive statistics for those variables with that of the original study.

On average, teachers increased in their scaffolding practices from September to April (*M*=1.48; mean of the increase). Teachers also improved in their fidelity of implementation over time (see Table II.2), which is unsurprising given that the fidelity of implementation measure is primarily focused on teachers' scaffolding. Teachers scored a 2.97 out of 5 on fidelity of implementation across the intervention period (monthly; 8 times). Across the four components, average fidelity for Buddy reading was 3.26, Play planning was 3.10, Make-believe play was 2.96, and Make-believe play practice was 2.62. On average, teachers gained .74 points in their fidelity scores across the intervention period.

For initial fidelity, out of a total 76 teachers who participated in the Tools of the Mind intervention, 30 teachers scored above a 3 on fidelity for the Buddy reading component of the intervention in the *first two months* of implementation (September/October). For Make-believe play, 19 teachers scored above a 3 in those first few months. For Play planning, 15 teachers scored above a 3. And for Make-believe play practice, 10 teachers score above a 3.

At the end of the intervention (March/April), 62 teachers scored above a 3 on Buddy reading, 41 teachers scored above a 3 on Play planning, 40 teachers scored above a 3 on Make-believe play, and 28 teachers scored above a 3 on Make-believe play practice. At the end of the intervention, only 30 Tools' teachers averaged a 4 or above across all four components.

Correlations

Pairwise correlations between study variables within the Tools group are presented in Table II.3. Teachers' average fidelity of implementation was not correlated with children's baseline or follow-up math or EF skills (child variables aggregated to the classroom level). Initial levels of fidelity for Play planning was statistically significantly correlated with baseline (r= .24) and follow-up (r= .37) math skills, but not EF skills (r's= .03 - .10). And initial Make-believe play and Make-believe play practice were statistically significantly correlated with baseline math (r=.14) and follow-up math (r=.11), respectively (but similar to Play planning, were less strongly correlated with EF measures (r's= .01 -.10)). Initial Buddy reading was statistically significantly correlated with baseline and follow-up math (r=.30, r=.33), as well as baseline and follow-up

Pencil Tap scores (*r*=.26, *r*=.26), but not with the other EF task (Head-to-Toes *r*'s=.07, .06). In contrast to initial fidelity of implementation, end fidelity of implementation across intervention components were generally negatively correlated with children's baseline and follow-up scores. End Play planning was negatively correlated with children's baseline and follow-up math skills (*r*'s = -.19, -.16), and baseline EF skills (*r*'s=- .21, -.07) (but not follow-up EF skills). End Make-believe play and end Make-believe play practice showed a similar pattern of significance and magnitude. End Buddy reading also displayed a similar pattern of significance, but magnitudes were smaller (*r*'s= -.06-.09).

For teacher characteristics, teachers' years of teaching preschool (*r*=-.06) and years of education (*r*= .09) were not statistically significantly correlated with average fidelity of implementation. However, teacher education was statistically significantly correlated with initial levels of fidelity of implementation across the four components (range *r*'s= .15 - .41), and teaching experience was not consistently correlated with initial levels of fidelity of implementation (*r*'s=.05- .12). For end fidelity, teaching experience was statistically significantly correlated with end fidelity of Make-believe play and Make-believe play practice (*r*'s= .12, .15), and teacher education was statistically significantly correlated with Make-believe play practice (*r*=.16).

For correlations listed thus far, I only analyzed the Tools condition (because there were only fidelity data for the Tools condition). However, for correlations with children's baseline and follow-up skills and teacher characteristics, there is a separate correlation table for the Control group (Table II.4). For the Tools group, teachers' years of experience was not statistically significantly associated with children's baseline or follow-up math and EF skills (range *r*'s=-.09-.04). However, for the Control group, teachers' years of experience was consistently statistically significantly associated with both children's baseline and follow-up math and EF skills (range *r*'s = -.24, -.31). For the Tools group, teacher education was statistically significantly correlated with children's math skills at follow-up only (*r*=.22). This pattern also emerged in the Control group, but the magnitude of the association was lower (*r*=.10).

Research Question 1: Replication Analysis

I first replicated the causal analysis of the Tools' impact on children's EF and math skills (see Table II.5 for an overview, Appendix A Table 4 for detailed results table). Tools did not have a statistically significant impact on children's gains in the Head-to-Toes Task (*ß* =-.004, *p*=.88), or Pencil Tap task (*ß* =-.003, *p*=.90), but it did have a trending towards statistically significant impact on children' math skills (*ß* =.12, *p*=.08). Children in Tools classrooms scored an average of 2.79 points (W-Score) more than children in the Control group at follow-up, which represents an effect size of .09 (Cohen's d). This replicated the original study's findings.

Research Question 2: Fidelity of Implementation Predicting EF and Math

In this section, none of these analyses were reported in the original full report of the Head Start CARES analyses (and to my knowledge, have never been conducted before). Average fidelity of implementation across the year overall and across specific component did not statistically significantly predict children's gains in EF or math skills (see Table II.6 for a summary, Appendix A Tables 5-8 for detailed results' tables). For initial fidelity of implementation (September/October) on each of the components, teachers' score on Play planning in the first month of implementation statistically significantly predicted children's math skills (*ß* =.21, *p*=.005); children whose teachers' achieved higher fidelity of implementation in Play planning gained 3.87 points more in math than children whose teachers' took longer to achieve higher fidelity on this component. This phenomenon did not emerge for the other three components of Tools' (see Table II.5), or for the EF outcomes. For end fidelity of implementation (March/April), only end fidelity score of Play planning statistically significantly negatively predicted children's gains in Pencil Tap ($ß$ =.12, p=.04).

Post-Hoc Analyses

As evidenced by the prior section, teachers who had initially high ratings on fidelity of implementation in the first month of implementation were more likely to have children who made more gains in math than teachers who took longer to achieve fidelity. To further characterize those teachers, I made the post-hoc decision to include Research Questions 3a and 3b.

Research Question 3a: Teacher characteristics and initial levels of fidelity of implementation. As evidenced by the correlations, teachers in the Tools of the Mind group who had a Bachelor's degree or higher (57% of teachers) and teachers with less than 10 years of teaching experience (26% of teachers) were more likely THAN WHO to achieve higher levels of fidelity of implementation in the first month of implementation on all four intervention components. However, this was not observed for average levels of fidelity. This suggests that teachers, regardless of education or experience, were likely to achieve similar average fidelity scores across the intervention period. In contrast, teachers with more education and less teaching experience achieved fidelity more quickly.

When examined closer, a total of 15 teachers (out of 76) scored above a 3 on fidelity of implementation in the first month of Play planning, and all 15 teachers had a Bachelor's degree or higher. Interestingly, of these 15 teachers, 10 of them had fidelity scores that remained relatively unchanged through April (changed < 1 point), four teachers decreased in their fidelity of implementation (>1 point), and one teacher increased in their fidelity of implementation (>1 point) from September to April.

Teacher characteristics predicting fidelity of implementation. Given the high correlations between teacher characteristics and fidelity (*r*'s= .15 - .41), I examined whether teacher education and teaching experience predicted initial high levels of fidelity of implementation (see Table II.7 for a summary). Teachers who had at least 10 years of experience (74% of teachers) did not have higher initial levels of fidelity of implementation on any component except Make-believe play practice (*ß* =.20, *p*=.001). However, teachers with a Bachelor's or higher degree (57% of teachers) statistically significantly predicted higher initial levels of fidelity of implementation in Play planning (*ß* =.24, *p*=.001), Make-believe play (*ß* =.17, *p*=.01), and Make-believe play practice (*ß* =.33, *p*=.001), but not Buddy reading (*ß* =.04, *p*=.28).

Research Question 3b: Tools' impact on math and EF by teacher characteristics. Considering that teacher education and experience statistically significantly predicted initial levels of fidelity on some components, and initial levels of fidelity also statistically significantly predicted children's math gains, I wanted to investigate whether teacher characteristics were driving this observed impact on math gains (see Table II.8 for an overview of results, Appendix A Tables 9-10 for detailed results tables). When the sample was restricted to teachers with no Bachelor's degree (*N*=33), teachers in the Tools group (compared to Control group) made 3.66 point gains more in math (*ß* =.11, *p*=.19), and .01 point fewer gains in Pencil Tap (*ß* =-.01, *p*=.79), and .03 more gains in Head-to-Toes task (*ß* =.01, *p*=.39). For teachers with a Bachelor's degree or higher (*N*=44), teachers in the Tools group (compared to Control group) had students who made .83 point gains more in math (*ß* =.08, *p*=.30), and .01 point gains more in Pencil Tap (*ß* =.03, *p*=.21), and .19 point gains more in Head-to-Toes task (*ß* = .01, *p*=.76). For teachers with 10 years or more of experience (*N*=57), teachers in the Tools group (compared to Control group) had students who made 1.44 more point gains in math (*ß* =.02, *p*=.49), .03 more point gains in Pencil Tap (*ß* =.04, *p*=.31), and 0.13 more gains in Head-to-Toes (*ß* =.01, *p*=.67). For teachers with less than 10 years of experience (*N*=20), teachers in the Tools group (compared to Control group) had students who made 1.06 more point gains in math (*ß* =.02, *p*=.72), .004 more gains in Pencil Tap (*ß* =.007 , *p*=.90), .24 less gains in Head-to-Toes (*ß* =-.02, *p*=.65).

I then fit combined models with a treatment indicator and teacher education as main effects, and an interaction between the two (also Table II.8; detailed results in Appendix A Table 11). The interaction was not statistically significant in predicting children's gains in math (γ = 2.16, *SE* = 1.49, *p* = 0.09, β = .12), or gains in EF (Pencil Tap: γ = 0.19, *SE* = 0.43, *p* = 0.65, β = 0.02; Head-to-Toes: γ = -0.14, *SE* = 0.52, *p* = 0.79, β = 0.004). For years of teaching experience, the interaction between teacher experience and Tools was not statistically significant in predicting children's gains in math (γ = -1.53, *SE* = 3.73, *p* = 0.63, β = -.01) or EF (Pencil Tap: γ = .03 *SE* = .04, *p* = 0.78, β = .003; Head-to-Toes: γ = .06, *SE* = .55, *p* = 0.29, β = .02),

In summary, subgroup analyses revealed that the impact of Tools on children's math gains was not driven by teachers' years of teaching experience. However, there was a trending towards statistically significant finding for teacher education. With a larger sample size, this subgroup finding may have reached conventional levels of statistical significance.

Discussion

In this chapter, I analyzed an existing EF intervention—the Tools of the Mind curriculum—to better understand the mechanism by which it had impacted preschoolers' math gains in the Head Start CARES evaluation (Morris et al., 2014). The two hypotheses I explored were whether differences in fidelity of implementation and differences between teachers explained the impact on children's math skills. One of the main motivations for examining this first hypothesis (fidelity of implementation) was based on the incongruency between the Tools of the Mind's theory of change and the fidelity tool used in the Head Start CARES study (and most evaluations of the Tools of the Mind curriculum). Teachers implementing Tools of the Mind should hypothetically reduce scaffolding interactions over time as children begin to internalize the different components of Tools and are able to carry out creating and following a "Play plan" on their own. However, teachers' scaffolding behaviors in the Tools of the Mind intervention in the Head Start Cares study, on average, increased over time. Yet, when analyzed further, there was substantial variation between teachers on how well they implemented the intervention, and how this fidelity changed over time. Due to this finding, my second hypothesis was that there was something unique about the teachers who implemented the intervention to fidelity in the beginning of the intervention (followed by a decrease in scaffolding techniques), in contrast to teachers who took longer to achieve fidelity, or stayed consistently "high" in their scaffolding practices across the intervention period. To explore this hypothesis, I examined whether differences in teachers' education and years of teaching experience explained impacts on children's math skills.

Fidelity of Implementation Hypothesis

Results revealed that teachers' average fidelity of implementation was not statistically significantly related to children's baseline or follow-up EF and math skills. However, teachers who scored higher on fidelity of implementation on two Tools' components – Play planning and Buddy reading—in the first month of implementation had children who scored higher on math at baseline and follow-up. This suggests a selection effect whereby teachers who were better implementers from the get-go also had children with initially higher math scores. In contrast, teachers who scored lower on fidelity of implementation at the end of the intervention on Play planning had children who scored lower on EF (Pencil Tap) at follow-up, but not baseline (e.g., no selection effect). This suggests that teachers who were still scaffolding the Play plan component of Tools at the end of the intervention had children with lower EF scores.

These statistically significant correlations laid the foundation for my investigation of how different approaches of measuring fidelity (average, baseline, and end) could account for differences in children's gains in math (and EF). Results revealed that average fidelity—overall and of specific intervention components—did not contribute to gains in children's skills. However, teachers who achieved high levels of fidelity on Play planning in the first two months had children who gained 3.97 points more on math at the end of the intervention compared to children in classrooms with low fidelity in the beginning of the intervention. This finding could be explained by the high baseline correlations between initial fidelity of implementation and children's math skills. Or perhaps there was something about high quality scaffolding of Play plans that transferred to children's capacity to learn from math instruction.

Teachers who exhibited high levels of fidelity at the end of the intervention on Play planning also had children who made fewer (.03-point) gains in EF (Pencil Tap measure). This finding aligns with prior work on how higher fidelity of implementation was associated with fewer gains in EF (Baron et al., 2020; Farran & Wilson, 2014). Fidelity observations in past evaluations also were collected in the latter half of the intervention period, which is close to the "end" fidelity time point of this study. This finding harkens back to an earlier point made in the introduction about how fidelity of implementation and specifically the scaffolding that is highlighted in this measure should be decreasing over time. Teachers who are still providing high quality scaffolding of Play plans at the end of the intervention may be a sign that children have not been given the opportunity to internalize this process on their own. Executive functioning can theoretically only be strengthened if children are given room to exercise it and rely less on teacher scaffolding.

This disconnect between fidelity of implementation and children's skills highlights a broader issue with the Tools of the Mind curriculum and theory of change. Unlike other successful preschool interventions, such as the Building Blocks curriculum, Tools does not prescribe teachers with a developmental approach when it comes to the targeted skill (e.g., EF). For example, in the Building Blocks curriculum, teachers are coached on how to begin with incremental scaffolding that helps children (for example) recognize shapes, and then identify shape attributes, and then discriminate shapes by attributes, and then recognize the shape only by the attributes (only touching shape not seeing shape). In EF interventions, like Tools, this developmental approach is more vague, and seen as more of a binary "can or cannot" self-regulate, rather than a continuum. The Tools of the Mind teaching manual does note how "teacher involvement should decrease over time as children begin to internalize turn-taking, planning, and role-playing skills" (p. 131, Bodrova & Leong, 1996), but clearly this is not reflected in the fidelity tool. In the Tools of the Mind curriculum, teachers are expected to scaffold until a child is able to execute an action on their own, but the steps of how to get there, or how to observe a child's "readiness" to reduce scaffolding, are ambiguous. Just as researchers have adapted a developmental approach to content-based curricula, it is important to consider this work with EF interventions as well.

Overall, this chapter highlights how more research is needed on the measurement of fidelity of implementation in EF interventions. There is suggestive evidence that the design of

EF interventions—and thus their fidelity of implementation measures—are not sensitive to the development of EF over time, and teacher practice should respond to this development. It is also important to note, however, that considering the number of models performed to examine this hypothesis (42), the above findings may also be spurious and should be interpreted with caution.

Teacher Characteristics Hypothesis

For my second hypothesis, I examined characteristics of the teachers who were achieving high levels of fidelity of implementation in the first few months (since initial fidelity of implementation was a statistically significant predictor of children's math gains). Correlations revealed that compared to average and change in fidelity, teachers who achieved higher initial levels of fidelity on two components of the Tools intervention were more likely to have less teaching experience (e.g., number of years teaching preschool), and teachers who achieved high initial levels of fidelity on all four components were more likely to have a Bachelor's degree or higher. Subgroup analyses by teaching experience and education subgroups, however, revealed no statistically significant impacts on children's math skills, although there was a trending towards significant interaction of teachers with a Bachelor's degree predicting greater gains in children's math skills in the Tools group.

This finding contributes to the policy debate in the literature on the qualifications needed for teachers to be effective implementers of interventions (Barnett & Frede, 2010). For example, there has been a growing push to ensure that all preschool teachers have a Bachelor's degree (Clarke-Stewart & Allhusen, 2005). And there is some evidence that teachers with a Bachelor's degree (but not more years of teaching experience) are better able to handle mixed-age classrooms (Ansari & Pianta, 2019). Yet others have argued that it is not the educational degree of the teacher, but the specific knowledge or courses taken that play a critical role in a child's development (Whitebook & Sharon, 2011), and how a comprehensive professional development experience could "provide the knowledge, skills and supports for teachers" (Early et al., 2007, p.577) to provide high-quality education even without a Bachelor's degree. Similarly, some studies have demonstrated that differences in fidelity of implementation by teacher education only emerge if there is complexity to a curriculum (Lieber et al., 2009; LoCasale-Crouch et al., 2007). Considering the complexity of Tools of the Mind—with over 60 activities and several changes to the schedule on a day-to-day basis—it is not surprising that there were differences in initial fidelity of implementation by teacher education. Future studies would benefit from collecting data on teacher education, knowledge, and fidelity of implementation to help disentangle these factors.

Improving Investigations of the "Black Box" of EF Interventions

Evaluations of interventions are helpful not only if they tell researchers if they work but why they worked. To date, the field knows that Tools of the Mind works for some children, but not others, and impacts some skills, but not others. But before we can definitively conclude the benefits of an early intervention, it is important to look beyond fidelity of implementation and examine how coaching and new curricula contributes to teachers' perception and understanding of the theory of change. For example, in the Head Start CARES study, there were no data on how well the teachers understood the theory of change of Tools beyond the coach survey question of whether "teacher(s) asked questions about the curriculum." Yet, several preschool studies have stressed the importance of teacher knowledge in predicting intervention success (e.g., statistically significant impacts on child gains; Cunningham, Zibulsky, & Callahan, 2009; Hamre et al., 2012; Spear-Swerling & Cheesman, 2011) and how this knowledge is not always related to observed fidelity of implementation (Pianta et al., 2012; Piasta, Justice, McGinty, Mashburn & Slocum, 2015). In elementary school studies, it is very common to assess teachers' understanding of the curriculum and compare and contrast this with teacher practice (Lakin & Shannon, 2015), yet this approach is not typically adopted in early childhood education research. Additionally, more data are needed by intervention researchers to better understand intervention impacts. For example, in the current study, there were no data available on teachers' classroom practices beyond related intervention components (e.g., general scaffolding). Intensive coaching of the Tools curriculum that is completely changing a teachers' schedule is bound to bleed into other areas of the classroom—like math teaching—which fidelity measures may not be able to pick up on. In the current study, teachers were asked how much they focused on academics compared to other areas, and only 6 out of 77 teachers in the Tools group reported a strong focus on this area at baseline. Yet, this was the only other data point that overlapped with any potential differences in a children's classroom math experiences. If the field is to do a better job at pinpointing the active ingredients and deciphering the black box, we need to include more measures of teacher understanding in our evaluations.

Limitations

There are several limitations of this study. First, the number of models fit may have led to spurious outcomes. Second, the sample size for the subgroup analysis by teacher characteristics was small with a power level of .42, thus and a larger sample size may have been better equipped to detect an effect.

Conclusion

In summary, this chapter did not provide answers on how the Tools of the Mind EF-intervention impacted children's math skills, but it did provide insight on three critical issues with EF-only interventions. One, data are lacking on teacher understanding of the theory of change of EF intervention, and how teachers may be applying “EF content” and learned practices to other areas of teaching. Two, similar to how past research has highlighted that EF interventions appear to only help children with the lowest EF skills, this chapter illustrated that EF interventions also appear to be more promising with some teachers than others. Future research needs to consider these two issues in evaluation of EF interventions by collecting more data on teachers specifically, which will also from a policy perspective contribute to our understanding of long-term sustainability of these interventions and factors influencing teacher buy-in. A final issue is that this chapter replicated a fundamental issue with the fidelity of implementation measure of the Tools of the Mind intervention. The lack of sensitivity of this measure over time points to a broader lack of understanding of how “EF supports” for children should change over time and how to detect changes in children’s EF development and respond accordingly. In a later chapter I address some of these issues by collecting data on teacher understanding of EF and examining differences between teachers and their fidelity of implementation.

Before further exploration of how these identified differences between teachers affects fidelity of implementation and intervention success, in the next chapter I provide pilot data for a different intervention approach—combining EF and math content. After presenting findings from this combined approach, in Chapter IV I then return to the issues raised in this chapter on teacher-related sources of variation in intervention fidelity.

References

Abry, T., Hulleman, C.D., & Rimm-Kaufman, S.E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation, 36(*3).

Ansari, A., & Pianta, R.C. (2019). Teacher-child interaction quality as a function of classroom age diversity and teachers' beliefs and qualifications. *Applied Developmental Science, 23*(3), 294-304.

Barnett, W.S., & Frede, E. (2010). The promise of preschool: Why we need early education for all. *American Educator.*

Barnett, W.S., Jung, K., Yarosz, D.J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effets of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*(3), 299-313.

Baron, A., Malmberg, L., Evangelou, M., Nesbitt, K., & Farran, D. (2020). The Play's the Thing: Associations between Make-Believe Play and Self-Regulation in the Tools of the Mind Early Childhood Curriculum. *Early Education and Development, 31*(1) 66-83.

Berk, L.E. (1992). *Children's Private Speech: An Overview of Theory and the Status of Research* in (Eds). Private Speech: From Social Interaction to Self-Regulation: Illinois

Bickman, L., Riemer, M., Brown, J.L., Jones, S.M…Massetti, G. (2009). Approaches to measuring implementation fidelity in school-based program evaluations. *Journal of Research in Character Education, 7*(2), 75-101.

Blair, C., & Razza, R.P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663.

Blair, C., McKinnon, R.D., & Daneri, P.M. (2018). Effect of the tools of the mind kindergarten program on children's social and emotional development. *Early Childhood Research Quarterly, 43*(2), 52-61.

Bodrova, E., & Leong, D.J. (2007). Play and Early Literacy: A Vygotskian Approach. From Play in Literacy in Early Childhood: Research From Multiple Perspectives.

Bodrova, E., & Leong, D.J. (2018). Tools of the Mind: A Vygotskian Early Childhood Curriculum. *International Handbook of Early Childhood Education*, 1095-1111.

Brown, E.T., Molfese, V.J., & Molfese, P. (2008). Preschool students learning in literacy and mathematics: Impact of teacher experience, qualification, and beliefs. *Journal of Education for Students Placed at Risk (JESPAR), 13*(1), 106-126.

Bueno, M., Darling-Hammond, L., & Gonzales, D. (2010). A matter of degrees: Preparing teachers for the Pre-K classroom. Education Reform Series. Pew Center on the States.

Cane, J., O'Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behavior change and implementation research. *Implementation Science, 7.*

Clarke-Stewart, A., & Allhusen, V.D. (2005). *What We Know about Childcare*. Harvard University Press.

Clements, D. H., Sarama, J., Unlu, F., & Layzer, C. (2012, March). *The efficacy of an intervention synthesizing scaffolding designed to promote self-regulation with an early mathematics curriculum: Effects on executive function.* Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Cohen, L.E., Kramer-Vida, L., Frye, N., & Andreou, M. (2014). The effect of bilingual instruction and play on preschoolers' English proficiency. *International Journal of Play, 3*(1)- 36-52.

Collins, L.M., Murphy, S.A., Bierman, K.L. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science, 5*, 185-196.

Cunningham, A.E., Zibulsky, J., & Callahan, M.D. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing, 22*, 487-510.

Daneri, M.P., Blair, C., Hammer, C., & Lopez, L. (2017, April*). Effect of the tools of the mind prekindergarten program on the school readiness of English language learners in head start*. Paper presented at the biennial meeting for the Society for Research in Child Development, Austin, TX.

Diamond, A., Barnett, W.S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387-1388.

Diamond, A., Lee, C., Senften, P., Lam, A., & Abbott, D. (2019). Randomized control trial of Tools of the Mind: Marked benefits to kindergarten children and their teachers. *PloS One, 14*(9)

Durlak J.A. (2008). Common risk and protective factors in successful prevention programs. *American Journal of Orthopsychiatry, 68*(4), 512-520.

Durlak, J.A., & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(327).

Early, D.M., Maxwell, K.L., Burchinal, M., Alva, S…& Zill, N. (2007). Teachers' Education, Classroom Quality, and Young Children's Academic Skills: Results From Seven Studies of Preschool Programs. *Child Development, 78*(2), 558-580.

Farran, D.C., & Wilson, S.J. (2014). *Achievement and self-regulation in pre-kindergarten classrooms: Effects of the Tools of the Mind curriculum*. Peabody Research Institute: Nashville, TN.

Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., Wallace, F…Shern, D. (2005). Implementation research: A synthesis of the literature.

Greenberg, M.T. (2010). School-based prevention: Current status and future challenges. *Effective Education, 2*(1), 27-52.

Gresham, F.M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38*, 254-272.

Hamre, B.K., Pianta, R.C., Burchinal, M… Scott-Little (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88-123.

Hildebrandt, S.A., & Eom, M. (2011). Teacher professionalization: Motivational factors and the influence of age. *Teaching and Teacher Education, 27*(2), 416-423.

Hulleman, C.S., & Cordray, D.S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Research on Educational Effectiveness, 2*(1) 88-110.

Humphrey, N., Barlow, A., & Lendrum, A. (2018). Quality matters: Implementation moderates student outcomes in the PATHS curriculum. *Prevention Science, 19*, 197-208.

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational*

Research, 85(4), 512-552.

Klimes-Dougan, B., August, G.J., Lee, C.Y.S., Realmuto, G.M., Bloomquist, M.L., Horowitz, J.L., Eisenberg, T. (2009). Practitioner and site characteristics that relate to fidelity of implementation: The Early Risers prevention program in a going-to-scale intervention trial. *Professional Psychology*: *Research and Practice, 40*(5), 467-475.

Lakin, J.M., & Shannon, D.M. (2015). The role of treatment acceptability, effectiveness, and understanding in treatment fidelity: Predicting implementation variation in a school science program. *Students in Educational Evaluation, 47*, 28-37.

Landry, S.H., Swank, P.R., Smith, K.E., Assei, M.A., & Gunnewig, S.B. (2006). Enhancing early literacy skills for preschool children: Bringing a professional development model to scale. *Journal of Learning Disabilities, 39*(4), 306-324.

Landry, S.H., Anthony, J.L., Swank, P.R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448-465.

Li, J., & Julian, M.M. (2012). Developmental relationships as the active relationships: A unifying working hypotheses of “what works” across intervention settings. *American Journal of Orthopsychiatry, 82*(2), 157-166.

Lieber, J., Butera, G., Hanson, M., Palmer, S., Horn, E…Odom, S. (2009). Factors that influence the implementation of a new preschool curriculum: Implications for professional development. *Early Education and Development, 20*(3), 456-481.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C…Barbarin, O. (2007). Observed classroom quality profiles in state0funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3-17.

Lonigan, C.J., & Phillips, B.M. (2012). *Comparing skills-focused and self-regulation focused academics and self-regulatory skills.* Society for Research on Educational Effectiveness

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–959

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology, 108*(1), 130–145.

Millaway, S.A. (2015). Impact of tools of the mind on middle school achievement.

Michie, S., Fixsen, D., Grimshaw, J.M., & Eccles, M.P. (2009). Specifying and reporting complex behavior change interventions: the need for a scientific method. *Implementation Science*, 4.

Molloy, L.E., Moore, J.E., Trail, J., Van Epps, J.J., & Hopfer, S. (2013). Understanding real-world implementation quality and “active ingredients” of a PBIS. *Prevention Science*, *14*, 593-605.

Morris, P., Mattera, S.K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact Findings from the Head Start CARES National Evaluation of Three Approaches to Improving Preschoolers’ Social and Emotional Competence. Executive Summary*. MDRC, New York, NY.

Mowbray, C.T., Holter, M.C., & Teague, G. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, *24*(3), 315-340.

Pianta, R.C., Hamre, B.K., & Allen, J.P. (2012). Teacher-students relationships and engagement: Conceptualizing, Measuring, and Improving the Capacity of Classroom Interactions. *Handbook of Research on Student Engagement,* 365-386.

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,* 731-755.

Ponitz, C.E.C., McClelland, M.M., Jewkes, A.M. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141-158.

Power, T.J., Bloom-Hoffman, J., Clarke, A.T., Riley-Tillman…Manz, P.H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools*, *42*, 495- 507.

Raver, C.C., Jones, S.M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's Impact on Low-Income Preschoolers' Preacademic Skills: Self-Regulation as a Mediating Mechanism. *Child Development, 82*(1), 362-378.

Solomon, T., Plamondon, A., O'Hara, A., Finch, H…Tannock, R. (2018). A cluster randomized-controlled trial of the impact of the Tools of the Mind curriculum on self-regulation in Canadian Preschoolers. *Frontiers in Psychology.*

Spear-Swerling, L., & Cheesman, E. (2011). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing, 25*, 1691-1723.

Telzrow, C.F., McNamara, K., & Hollinger, C.L. (2000). Fidelity of problem-solving implementation and relationship to student performance. *School Psychology Review, 29*(3), 443-461.

Tout, K., Zaslow, M., & Berry, D. (2006). *Quality and qualifications: Links between professional development and quality in early care and education settings*. Critical issues in early childhood professional development: Brookes Publishing Company: Baltimore, MD.

Vygotsky, L.S. (1987). *The development of scientific concepts in childhood*. Plenum New York.

Wasik, B.A., & Hindman, A.H. (2014). Understanding the active ingredients in an effective preschool vocabulary intervention: An exploratory study of teacher and child talk during book reading. *Early Education and Development, 25*(7), 1035-1056.

Woodcock, R.W., McGrew, N., & Mather, K.S. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Riverside Publishing Company.

Table II.1. Overview of Fidelity components in the Tools of the Mind curriculum

Fidelity components of Tools: **1.) Play planning (Plan):** Teachers support all students while developing their play plans. A play plan might include the child's name, an illustration of the plan showing what the child will do during play time, lines representing the words of the play plan, and starting/beginning sounds to each word. **2.) Make-believe play (MBP)** Teachers set aside a required block of time for play with a clear theme. Teachers circulate the room scaffolding play and encouraging deep child engagement in play. **3.) Make-believe play practice (MBPP)** Teachers create inventive and engaging scenarios that are meant to be carried over by the children into their own play. This activity is meant to have a playful feel with high levels of student participation. **4.) Buddy reading (BR)** Teachers support students in engaging in turn-taking with peer buddy reading

Table II.2. Fidelity of Implementation on Specific Components of Tools (*N*=76 teachers)

	Initial Fidelity	End Fidelity	Average
Play planning	*M*= 2.19 [1, 5]	*M*=3.54 [1, 5]	*M*=3.15[1.21, 4.88]
Make-believe play	*M*=2.32 [1, 5]	*M*=3.29 [1, 5]	*M*=3.07 [1, 4.88]
Make-believe play practice	*M*=1.92 [1, 4]	*M*=2.89 [1, 5]	*M*=2.29 [1, 4.33]
Buddy reading	*M*=2.53 [1, 4]	*M*=3.87 [1, 5]	*M*=3.37 [1.29, 4.88]

Table II.3. Pairwise Correlations Between Study Variables Within Treatment Group

	Tools of the Mind Intervention															
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Child Variables (Aggregated to classroom-level)																
(1) Pre WJAP	1.00															
(2) Post WJAP	.83***	1.00														
(3) Pre PT	.57***	.53***	1.00													
(4) Post PT	.43***	.49***	.51***	1.00												
(5) Pre HT	.46***	.45***	.40***	.29***	1.00											
(6) Post HT	.58***	.60***	.51***	.48***	.49***	1.00										
Teacher Characteristics																
(7) Teacher Ed.	-.02	.22***	-.02	.08	.12**	.04	1.00									
(8) Teacher Exp.	.04	.04	-.01	-.08	-.09	-.08	-.22***	1.00								
Fidelity of Implementation																
(9) Avg Fidelity	.02	-.02	.03	.01	.00	-.02	-.06	.09	1.00							
(10) Int Play	.24***	.37***	.05	.07	.03	.10*	.37***	.05	.52***	1.00						
(11) Int MBP	.14**	.11**	.01	.05	.08*	.08*	.15***	-.12**	.61***	.78***	1.00					
(12) Int MBPP	.03	.11**	-.00	.05	.01	.10**	.41***	.05	.47***	.54***	.68***	1.00				
(13) Int BR	.30***	.33***	.26***	.26***	.07*	.06	.25***	-.12**	.57***	.54***	.51***	.63***	1.00			
(14) End Play	-.19**	-.16**	-.02	-.21***	-.07*	-.06	-.01	.02	.60***	.27***	.31***	.24***	.07*	1.00		
(15) End MBP	-.17*	-.19**	-.25***	-.12**	-.07*	-.11***	.05	.08*	.60***	.14***	.23***	.32***	.31***	.67***	1.00	
(16) End MBPP	-.27**	-.15**	-.31***	-.22**	-.15**	-.06*	.16***	.15**	.53***	.03	.07	.35***	.18**	.61***	.79***	1.00
(17) End BR	-.06*	-.09**	-.06*	-.01	-.08*	.07*	-.07*	.12**	.69***	.16**	.21***	.12**	.12**	.73***	.72***	.56***

Note. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Play= Plan planning. *p<.05 ** p<.01 ***p<.001

Table II.4 Pairwise Correlations Between Study Variables Within the Control Group

	Control Group							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Child Variables (Aggregated to classroom-level)								
(1) Pre WJAP	1.00							
(2) Post WJAP	.88***	1.00						
(3) Pre PT	.66***	.63***	1.00					
(4) Post PT	.50***	.49***	.57***	1.00				
(5) Pre HT	.62***	.57***	.57***	.39***	1.00			
(6) Post HT	.60***	.65***	.65***	.58***	.60***	1.00		
Teacher Characteristics								
(7) Teacher Ed.	-.06	.11*	.06	.11*	.11**	.04	1.00	
(8) Teacher Exp.	-.26**	-.24**	-.25**	-.31**	-.31**	-.25**	-.12*	1.00

Note. *p<.05 ** p<.01 ***p<.001

Table II.5. Research Question 1: Impact of Tools of the Mind on Children's Math and EF Gains (Replication Analysis)

	WJAP	Head-to-Toes	Pencil Tap
Tools Overall Impact (Replication)	2.79 (1.48) +	-.04(.25)	-.00 (.02)

Note. Each model controls for child age, baseline WAP score, baseline expressive language score, and a flag for whether the baseline score was imputed or not. Random intercepts are at the center and classroom level and there is a fixed effect for block. This model is identical to the one used in the original Head Start CARES analyses. $+p<.10$ $*p<.05$ $**\ p<.01$ $***p<.001$

Table II.6. Research Question 2: Unstandardized Coefficients For Each Fidelity Measurement Approach Predicting Math and EF Gains

	WJAP	Head-to-Toes	Pencil Tap
Average Fidelity (All)	-.87(1.23)	.15(.19)	.01(.01)
Change Fidelity (All)	.03(1.10)	.21(.17)	-.00(.01)
Average Fidelity- BR	-.89(1.28)	.15(.20)	.01(.01)
Initial Fidelity- BR	1.78(1.41)	.16(.21)	.01(.01)
End Fidelity- BR	-.83(1.01)	.11(.17)	.01 (.01)
Average Fidelity- MBP	-.85(1.11)	.15(.19)	.00(.01)
Initial Fidelity- MBP	-.02(1.32)	.24(.19)	.01(.01)
End Fidelity- MBP	-.90(.90)	.22(.14)	.01 (.01)
Average Fidelity- MBPP	.44(1.14)	.18(.20)	.00(.01)
Initial Fidelity- MBPP	2.52(1.62)	.31(.20)	.01(.01)
End Fidelity- MBPP	.94(.79)	.13(.14)	.01(.01)
Average Fidelity-Plan	-.30(1.30)	.21(.04)	.01(.01)
Initial Fidelity- Plan	3.97(1.13)**	.18(.18)	.00(.01)
End Fidelity- Plan	-.03 (.98)	-.15(.16)	-.03 (.01)*

Note. Each cell is an unstandardized coefficient for a regression model (42 total models). Each model controls for child age, baseline WAP score, baseline expressive vocabulary, and flags for whether the baseline score was imputed or not. Random intercepts at the center level. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Plan = Plan planning. Initial fidelity was calculated averaging September and October scores, and End fidelity was calculated averaging March and April scores. +p<.10 *p<.05 ** p<.01 ***p<.001

Table II.7. Research Question 3a: Teacher Characteristics Within Tools Group Predicting Fidelity of Implementation

	Initial Fidelity-BR	Initial Fidelity-MBP	Initial Fidelity-MBPP	Initial Fidelity-Play
Teacher has a BA	.07(.07)	.35(.07)***	.51(.06)***	.57(.05)***
Teaching Experience 10+ years	-.12(.07)+	.11(.07)	.37(.06)***	.04(.06)

Note. Each cell is an unstandardized coefficient (8 total models) Random intercepts at the center (*N*=26) level. +*p*<.10 *p<.05 ** *p*<.01 ***p<.001

Table II.8. Research Question 3b: Treatment Impacts by Teacher Education and Teaching Experience Subgroups

	WJAP	Head-to-Toes	Pencil Tap
Teacher education main effect (restricted)			
Teachers without a BA	2.66(2.55)	.03(.13)	-.01(.03)
Teachers with a BA	.83(2.35)	.19(.33)	.01(.03)
Teacher experience main effect (restricted)			
Teachers < 10 years of experience	1.06(2.92)	-.24(.52)	.004(.04)
Teachers 10 years or more of experience	1.44(2.16)	.11(.34)	.03(.03)
Interactions (unrestricted)			
*Teacher BA*Treatment*	2.16(1.49)+	-.14(.52)	.02(.04)
*Teaching > 10 years * Treatment*	-1.53(3.73)	.06(.55)	.03(.04)

Note. Each cell is the unstandardized coefficient of the Tools' impact on math skills restricted to different teacher characteristics (12 total models). The unrestricted models (6 total models) include an interaction term for teacher and teaching experience (0/1). Random intercepts are at the classroom and center level and there is a fixed effect for Block ID. +p<.10 *p<.05 ** p<.01 ***p<.001

CHAPTER III: Feasibility of an Integrated Math and Executive Function Intervention with Preschoolers

This chapter explores the feasibility of explicitly integrating EF skills (e.g., working memory, response inhibition, cognitive flexibility) with math content (e.g., numeral recognition, addition/subtraction, patterning, counting-on, magnitude comparison, and shape recognition), with preschool children. I conducted this pilot study outside the classroom with small groups of children in accordance with Goal 2: Development and Innovation, of the IES recommendations for education research (IES, 2019). Results from this development phase inform my implementation phase in Chapter IV where I connect my findings from Chapter II to a combined math/EF intervention approach.

Theoretical Framework

I adopted a learning trajectories framework in the design of the math portion of my math/EF activities. For example, I included math domains that aligned with California Preschool Learning Foundations set of math standards: 1.) number sense & operations (e.g. reciting numbers in order, numeral recognition, one-to-one correspondence, cardinality, magnitude comparison, "counting on", addition/subtraction;2.) geometry (e.g. shape recognition, sorting/classifying based on shape attribute, composing shapes, spatial relations); 3.) patterning (e.g. extending and creating patterns; California Department of Education, 2008). Within each domain, Sarama and Clements (2009) have detailed a developmental progression For example, when counting, children have to be able to count objects in front of them (demonstrating one-to-one correspondence) before they can tell you that the last number designates how many objects in the group (demonstrating cardinality). Both skills are precursors to being able to *produce* a desired number of objects from scratch (can count out "5" objects).

The theoretical framework for integrating math learning trajectories *with EF* is rooted in incidental and implicit learning theory which describes how children are able to learn without explicit instruction, and instead engage in a non-conscious, spontaneous selective attention to acquire information. There are individual differences in children's implicit and incidental learning in early childhood, and these gaps in ability to spontaneously attend to relevant information reach stability beginning in middle childhood (Hagen & Hale, 1973). There is some work suggesting that incidental learning can be improved if some explicit supports are provided at first, and then reduced as children learn to spontaneously attend on their own. This phenomenon, however, has been primarily studied in literacy contexts and not when children are engaging in early math learning (Geis & Hall, 1976; Pelham & Ross, 1977). Implicit and incidental learning has also been linked to deficits in executive functioning (Barnes et al., 2010; Klinger et al., 2007), which highlights the importance of providing more explicit supports and instruction for children with lower EF skills.

Another theoretical underpinning motivating the combination of EF and math content is that most of mathematical thinking already involves cognitive flexibility, inhibition, and attention control. For example, children use cognitive flexibility when they are switching between different units (e.g., feet and inches) or different operations (e.g., addition and subtraction), as well as when they have to switch to a new solution strategy after another strategy has failed. One study illustrated that children's patterning ability—compared to other math skills— was related to cognitive flexibility in first grade (Bock et al., 2015). Children also have to engage in response inhibition when they have to suppress the first solution that comes to mind, or ignore salient extraneous information in a math word problem. Children also employ working memory since a majority of math tasks involve maintaining, manipulating, and adding relevant information while simultaneously engaging in another cognitively demanding task. A specific math example is when children are doing a series of operations in their head (e.g., mental math; “You have 5 pencils…now add 2… now take away 3” etc.). Past work has shown that children’s visual-spatial working memory predicted children’s ability to make magnitude judgements and engage in addition (Simmons, Willis, & Adams, 2012). Considering the theoretical framework underlying EF and math integration, this provides reasonable rationale for combining the two skills in my proposed EF/math activities.

Research Aims

In the current study, I tested an integrated math/EF intervention in which children explicitly exercised EF skills during math games. My research aims are as follows:

1) What is the feasibility of integrating EF and math content with preschool-aged children (e.g., are children engaged and do they progress through the different levels of EF demands and math skills?)

2) What is the impact of integrated EF/math activities on preschoolers’ EF and math skills?

Method

I recruited one public preschool center with four classrooms located in a suburban area of Southern California. I sent home letters to all 4-year-old children in this preschool center and parents of 64 children signed consent forms for their child to participate in the study (out of a total eligible 83). Four children were excluded from the study because of either uncooperative behavior or because they spoke no English, for a total sample of *N*=60 children.

Baseline Measures

Before randomization, a trained researcher and I tested all children (*N*=60) individually on math and EF skills in an empty classroom at the preschool center (when children from that classroom were outdoors).

Executive functioning. I assessed executive functioning using the Head-Toes-Knees-Shoulders (HTKS) Task (Cameron Ponitz et al., 2008). In this assessment, children are first told to touch their head (or toes) when asked to touch their toes (or head), and then asked to touch their knees (or shoulders). The rules are then switched, with head and knees paired with shoulders and toes. The task requires working memory to hold the rules in mind, inhibitory control to execute the opposite action, and attention shifting when the rules change. This task has been validated as a measure of behavioral self-regulation in multiple studies (Cameron Ponitz et al., 2009; McClelland et al., 2014) and has demonstrated strong internal and inter-rated reliability with diverse populations of children (McClelland et al., 2014).

Math. To assess children's math skills, I used the Applied Problems subtest of the Woodcock Johnson-III Tests of Achievement (Woodcock, McGrew, & Mather, 2001). This subtest includes questions assessing children's understanding of quantities and simple calculations that progressively get more challenging. This subtest has been used extensively and has been validated in previous work (McGrew & Woodcock, 2001).

Demographic information. As part of the parent consent form, parents filled out a short demographic survey reporting on their child's age and gender, as well as their family's household income. These variables were used as covariates in analyses.

Random Assignment

One classroom was randomly chosen as an all "treatment" classroom because one classroom was needed to pilot circle-time activities, which would have been difficult to orchestrate if there were control children in the classroom (e.g., *N*=20 treatment children in Classroom 1). Children in the remaining three classrooms were matched on math performance (WJ Applied Problems score), then randomly assigned to treatment or control groups within classrooms (Classroom 2, *N*=7 treatment, *N*=7 control; Classroom 3, *N*=4, *N*=4; Classroom 4, *N*=9, *N*=9). Only the randomized classrooms (*N*=20 treatment; *N*=20 control) were used to evaluate the impact of the activities on children's math and EF gains.

Procedure

An overview of the study procedure is in Figure III.1. Children within the treatment group were divided into smaller groups of two to three children by math skill level, specifically based on their knowledge of cardinality. This math skill aligned the most with knowledge needed for certain levels of the math/EF activities (low=does not understand cardinality, *N*=14; mid= understand cardinality but cannot add 1, *N*=16; high=understand cardinality and can add 1, *N*=10). After forming groups, another researcher and I took children to an empty classroom (when children from that classroom were outdoors) in pairs to practice EF/math activities at least three days per week, 15 minutes each day, for a total of four weeks. Every week, three integrated EF and math activities were introduced, and activities from the prior week were repeated. Within each session, children played an average of three activities (e.g., two activities determined by the researcher, one activity chosen by the children). Children in the treatment group were pulled out of their classroom during either free-choice/center-time or outdoor play time. Thus, children in the treatment group did not miss out on the designated Creative Curriculum whole-class "math" large-group that occurred 1-2 days per week in each classroom. However, because children were sometimes taken out of the classroom during center-time, they may have missed out on math content at math centers.

In order to determine whether these activities would work in a classroom context, in the remaining three weeks, I introduced activities during circle time (in the treatment classroom with 20 children) and center-time (for the remaining three classrooms that had between 5-8 children in the treatment and control groups each). In the three classrooms where center-time was implemented, children stayed in their same math skill groups, and the math center was "closed" to children in the control group who had to instead choose other activities for that day. Every session outside of the classroom was videotaped, but sessions inside the classroom were not video-taped.

Children in the control group received "business as usual" which included 1-2 large-group math activities per week that focused on either shape recognition or number recognition (these were the only math large-group activities observed during the 7 weeks of the intervention). Children in the control group also had the opportunity to visit the math center every day. At the math center, there were typically different materials available (e.g., geo boards with rubber bands, counters, number matching games, building materials e.g., Magna tiles) but there was no structured activity for children to follow.

Math/EF Activities

The intervention consisted of 10 activities that spanned a range of age-appropriate preschool math skills—geometry, counting, adding/subtracting, magnitude comparison, cardinality, one-to-one correspondence, and patterning. Each activity explicitly exercised children's EF, with some activities highlighting one EF skill (working memory, cognitive flexibility, or response inhibition) and others a balanced contribution of each skill. For example, one of the activities required children to identify a number and just name the number (if the number was red) or say which number was "one more" (if the number was blue). The game then proceeded by switching the rules and the red number then indicated "one more"—e.g., a rule switch which recruits the cognitive flexibility component of EF. Another level of the game included other colored numbers with rules such as "two more" or "one less," which further taxed working memory. Other activities placed more emphasis on EF, rather than practicing a math skill. For example, one of the activities involved children walking around in a circle on the carpet, during which they had to "go" and "stop" given a certain number rule (e.g., "when you hear '1', 'stop', all other numbers mean 'go'"). An overview of the 10 games and their corresponding developmental progression levels are presented in Table III.1.

Feasibility Measure(s)

To test feasibility of the activities, I first assessed child interest and engagement in the games. To do this, another researcher and I re-watched the videos and noted whether children expressed interest in continuing to play the games and whether they became disengaged or restless. To quantify this, in Table III.3, "Most popular" indicated how many children (of the total 40) requested to play that game at least once during the intervention. "Least popular" indicated how many children requested to end a game early or became visibly frustrated so I terminated the game and moved on to another activity.

In addition to child interest, I observed individual children's progression through the games. This was of key importance for this study because if the activities were too easy or too hard, then I would need to adjust accordingly in future iterations of this intervention. To quantify this, I noted whether a child "passed a level" when they were able to answer questions correctly at a given level for a consecutive three trials. Every time children returned to that game, they participated in three trials at each level prior to the level they were currently engaging in. This was to ensure that children had not regressed after time away from a given activity. Children were noted as "progressing through at least one level" during the intervention (Column 4 of Table III.2) if they initially did not pass a level but then achieved that level and moved onto the next level. Children who passed level 1 their first time and moved straight to level 2—but never mastered level 2—were noted as *not* progressing through a level for that activity.

Analytic Design

To assess whether children in the treatment group differed from children in the control group, I conducted balance checks to test for any statistically significant differences between gender, household income, baseline EF, or baseline math skills, as well as an overall F-test. As evidenced by Table III.2, there were no statistically significant differences between the treatment and control groups.

To examine the impact of the intervention on children's EF and math skills I fit a regression model with each of the outcomes (HTKS and WJ Applied Problem Raw score) regressed on a treatment dummy (1= treatment, 0= control), controlling for baseline EF/math, and a vector of child-level covariates (e.g., age, gender, and household income). I also included a fixed effect for block (child pairs matched on math ability within each classroom). There were no missing data for any of the variables of interest in the model, and there was no attrition.

Post-hoc, I also decided to examine differences in children's math and EF gains by format of the activities to see if some contexts (e.g., pairs vs. center time) were accounting for the impacts on children's EF skills. . I used the same covariates as in the first regression model.

Results

On average, treatment children spent 414 minutes participating in small-group math activities with me outside of the classroom (range 284-501 minutes, *SD*= 191), and an average of 180 minute in center time. Center times were estimated and not crosschecked with video footage because these sessions were not videotaped.

Feasibility

Child interest. The most popular activities (e.g., activity requested most from children at the end of each session) were the Clapping Game and the Number Freeze Game. These games were requested at least once by 34 and 39 children (out of the 40), respectively. The least popular games (most often terminated early due to child boredom) were the Shape Search, Number Share, and Pattern Match games. One of the most notable barriers in engagement of these activities was for children of the lowest math and EF skill levels. Nearly half (*N*=14) of the intervention group had difficulty recognizing numerals, understanding cardinality and recognizing shapes. This difficulty proved to be a barrier for most activities, so time spent with this subgroup of children was focused on either learning numbers, introducing shapes, or playing the "EF part" of the activity with little to no math emphasized (see Level 1 of "Color Cards" for an example).

Child progress. Across the 10 games, 30 children (out of 40 total treatment children) progressed through at least one level of one of the activities after seven weeks (see Table III.2), and 13 children mastered the highest level (level 3) across the 10 games. Specifically, children progressed through the most "levels" of basic number sense activities. For example, over half (*N*=23) of the children in the sample progressed at least *two* levels of the Toy Search ("counting on") and the Clapping Game ("one more", "one less") game. Conversely, more than a third of children could not even pass Level 1 of the (*N*=14; Shape Search) and shape attributes (*N*=13;

Shape Cards) games. Other "too difficult" activities were the magnitude comparison game (Number Share) and an activity where children had to remember how many counters where hidden in a cup while adding and subtracting counters to the cup (Magic Cups).

Impact of Intervention

Children's HTKS and WJ Applied problems scores were correlated at baseline (r=.39) and at follow-up (r=.56) for the whole sample. Interestingly, when correlational analyses were separated by control and treatment groups, children's baseline EF and math scores were similarly correlated (r=.37 control group, r=.40 treatment group), but at follow-up, math and EF scores in the control group remained similarly correlated (r=.40), whereas EF and math scores in the treatment group were more correlated at (r=.58).

After seven weeks, students in the randomized sample who participated in these activities (N=20) made statistically significant gains in executive functioning (HTKS assessment) compared to the control group (N=20; effect size= .17, p=.003), but there was no statistically reliable difference in math (measured by Woodcock Johnson Applied Problems subtest) between the two groups (effect size= .06, p= .12, see Table III.4). Overall, this demonstrated that, on average, children in the treatment group—who started with not statistically different EF scores from the control group—had EF scores that were 8.20 or 9.73 (depending on the sample) points higher at the end of 7 weeks than children in the control group.

Since there was variability in time children spent engaging with the intervention (due to attendance and child interest), I also examined whether time spent in different formats (pairs outside of classroom, centers, and circle time) predicted EF and math gains (Table III.5). I first conducted correlations to see whether children with higher or lower baseline EF and math skills spent more time in activities. Correlations revealed that children with higher baseline math scores did spend more time in pairs outside of the classroom (*r*=.44), but children with higher EF scores did not spend more time in pairs (*r*=.03). Regression analyses (controlling for age and gender) revealed that time spent in pairs statistically significantly predicted gains in EF (β=.32, *p*=.002), but not math (β=.01, *p*=.83). Furthermore, time spent in centers and circle time did not predict gains in EF (β=.10, *p*=.23; β=-.04, *p*=.73) nor math (β=-.07, *p*=.29; β=.70, *p*=.29).

Discussion

Overall, this pilot study utilized an implementation framework to study the potential of a set of activities to strengthen children's EF and math skills, and be adapted to the real-world setting of a classroom. This study is in the early or developing stage of implementation study where the primary aim is to refine the activities before evaluated at a larger scale (Hsueh, Halle & Maier, 2020). Results revealed that children were relatively engaged and interested in a majority of the developed math/EF activities. The most popular activities were embedded with movement, and involved practicing "counting on" and "one more" and "one less" math skills, with most children progressing through at least one level of these games. However, this study also revealed that some activities were too difficult or boring for children. For example, the shape games were too difficult for more than a third of children, and several children became bored in the games where there was little movement and less opportunity for "hands on" participation. This has implications for the revision of some of the math/EF activities and how children of the lowest math skill levels may not be developmentally ready to engage in this combined approach.

I also examined the impact of these math/EF activities on children's math and EF skills. Results illustrated that children in the full sample (40 treatment, 20 control) made statistically significant gains in executive functioning after seven weeks of playing the integrated EF and math content activities. In the matched-on-math-skill sample (20 treatment, 20 control), findings were similar. Relatedly, children in the treatment group had higher correlated math and EF scores than children in the control group post-intervention. Within the treatment group, children who spent the most time in pairs outside of the classroom made more gains in EF, and children with lower baseline EF skills but knowledge of cardinality also made statistically significant more gains in EF (compared to (1) children with low baseline EF/and no understanding of cardinality and (2) mid-high baseline EF children, regardless of their cardinality knowledge.

This combined math and EF intervention approach demonstrated a larger effect size on EF skills (d=.15) than EF-only interventions (d's <.10), but a smaller effect size on EF than math-only interventions (d's range .20 - .27). However, because I did not compare this combined approach to an EF-only or math-only group, I cannot conclude that this method is better or worse than an "only" approach.

More time spent engaging in the intervention (e.g., child-level dosage) also statistically significantly predicted EF gains. This highlights that dosage appears to matter for EF impact, and future work should consider "thresholds" of dosage that may be needed for children to maximally grow in EF skills. Yet, I also think that children spent more time in the intervention because they had a foundational knowledge of math and were more likely to remain engaged in the activities (and less likely to end the game early). This is an important piece to consider in future EF and math interventions in identifying which approaches or combinations of approaches work for which children.

Overall, the impacts on EF should be interpreted with caution, however, because it could be the case that children in the treatment group made specific gains in EF and not math because the HTKS assessment (that children were tested on pre and post) was very similar to many of the activities (e.g., implement opposite rule), so by the end of the intervention, the treatment group was familiar with this format. Another potential confound was that receiving extra adult-directed instruction and attention outside of the classroom may have improved children's EF skills regardless of the EF-content. An improved study design would include control arms in which children were taken out of the classroom to only practice math with no EF integration, and another group would practice EF with no math integration.

Unfortunately, this pilot study did not strengthen children's math skills. It is possible that children may have needed more than seven weeks to surpass their peers in the control group. Other math curricula that have documented statistically significant impacts on math range from 26 – 32 weeks of intervention dosage (Clements & Sarama, 2007; Presser et al., 2015) and they also used more developmental sensitive measures that cover a wider range of math skills (e.g., the REMA vs. Woodcock Johnson Applied Problems Subtest; Weiland et al., 2012). Children who had the lowest baseline math scores also happened to spend the least amount of time in the intervention, primarily because their sessions were terminated early due to disinterest . Thus, the children who had the most "room to grow" in their math skills were not "treated" as much as children who had "fewer math skills to gain." Children who spent more time in activities—and thus received "more" of the treatment—were also more likely to stay engaged in the activity, perhaps because they had a foundation of math skills that made it easier for them to progress from one level to the next and not lose motivation. I think this finding again confirms my conclusion that that participation and engagement in combined EF/math games can be tricky for lower-skilled learners.

Limitations

One limitation of this study was that some children spent more time in activities introduced at the beginning of the intervention because these activities were more likely to be repeated. It is possible that these beginning activities influenced the identified "Most popular" activities because child engagement could have been indistinguishable from familiarity and confidence with playing a game they had practiced multiple times. Additionally, some activities were identified as too boring or too difficult for children, and would likely even be less engaging when implemented more frequently in a large group setting by the teacher. Another limitation was that the small sample size weakened the ability to detect impacts on children's EF and math skills. As such, all impact results should be interpreted with caution. Lastly, the feasibility of math/EF activities were only assessed in the context of one preschool center, which limits the external validity of the study.

Conclusion

Overall, this chapter presented encouraging evidence that some combined math/EF activities are engaging for children, and most children improved in their EF skills across the intervention period. Specifically, the activities that had a movement element were the most popular among children, and children made the most improvement in activities where they practiced number sense skills. While the findings from the study are encouraging in the feasibility of a combined EF/math approach, there were also some limitations. The next chapter addresses some of these limitations and investigates whether a revised version of these math/EF activities can be implemented by teachers with high fidelity, supported by collaborative coaching.

References

Barnes, K.A., Howard, J.H., Howard, D.V., Kenealy, L., Vaidya, C.J. (2010). Two forms of implicit learning in children. *Developmental Neuropsychology, 35*(5), 494-505.

Institute for Education Sciences (IES) Goal Structure (2019). National Board for Education Sciences Annual Report.

Geis, M.F., & Hall, D.M. (1976). Encoding and incidental memory in children. *Journal of Experimental Child Psychology, 22*(1), 58-66.

Hagen, J.W., & Hale, G.A. (1973). The development of attention in children. *Educational Testing Service,* Princeton, N.J.

Kalra, P.B., Gabrieli, J.D.E., & Finn, A.S. (2019). Evidence of stable individual differences in implicit learning. *Cognition, 190*, 199-211.

Klinger, L.G., Klinger, M.R., & Phelig, R.L. (1999). Implicit Learning Impairments in Children.

Pelham, W.E., & Ross, A.O. (1977). Selective attention in children with reading problems: A developmental study of incidental learning. *Journal of Abnormal Child Psychology, 5*, 1-8.

Presser, A.L., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big Math for Little Kids: The Effectiveness of Preschool and Kindergarten Mathematics Curriculum. *Early Educational and Development, 26*(3) 399-426.

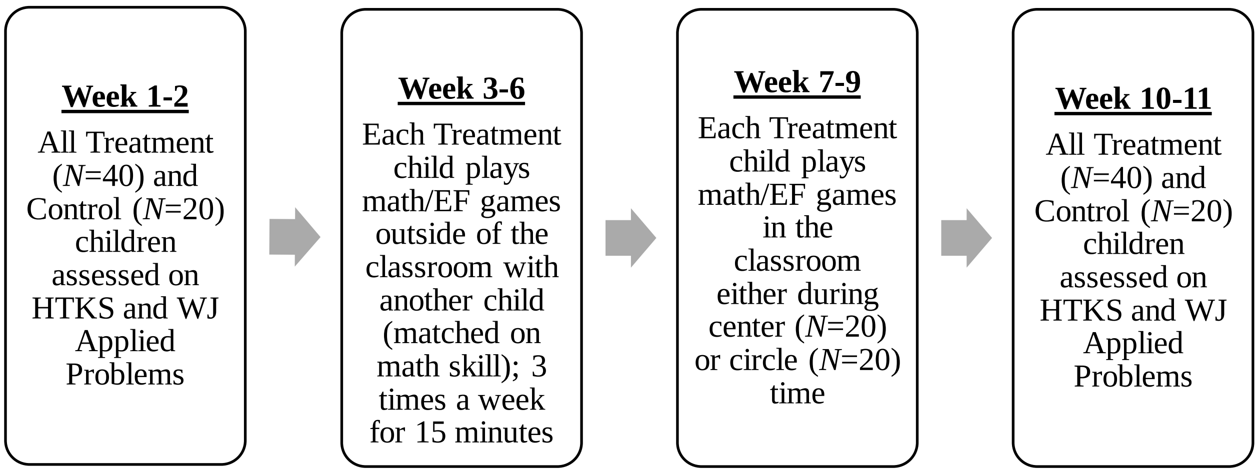
Ponitz, C.E.C., McClelland, M.M., Jewkes, A.M. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141-158.

Sasser, T.R., Bierman, K.L., Heinrichs, B., & Nix, R.L. (2017). Preschool intervention can promote sustained growth in the executive function skills of children exhibiting early deficits. *Developmental Science, 28*(12), 1719-1730

Tominey, S.L., & McClelland, M.M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education and Development, 22(*3), 489-519.

Woodcock, R.W., McGrew, N., & Mather, K.S. (2001). *Woodcock-Johnson III tests of cognitive abilities.* Riverside Publishing Company.

Figure III.1. Timeline of Study II Activities



Note. Control children received "business as usual" which included 1-2 large-group math activities per week that focused on either shape recognition or number recognition (these were the only math large-group activities observed during the seven weeks). Control children could also visit the math center every day where there were different materials available (e.g., geo boards with rubber bands, counters, number matching games, building materials e.g., Magna tiles) but there was never any structured activity (nor teacher present) for children to follow.

Table III.1. Math and EF Skills Required Across Each Level Within Each Activity

	Level 1	Level 2	Level 3
<u>Game 1: Color Cards</u> Math: Numeral recognition, operations EF: Working memory	Remember two directions with numbers 1-5 (or 1-10, depending)	Addition with 1 & remember two directions (1-10)	Addition with 1 and 2 & remember three directions (1-10)
<u>Game 2: Head, Tummy, Knees</u> Math: Numeral recognition EF: Working memory, cognitive flexibility	Memorize two numbers & two corresponding motoric movements	Memorize three numbers & three corresponding motoric movements Memorize two numbers & *switching* the corresponding motoric movement from Level 1	Memorize three numbers & *switching* the corresponding motoric movement from Level 2
<u>Game 3: Number Freeze</u> Math: Numeral recognition EF: Working memory, cognitive flexibility, inhibitory control	Memorize two numbers & two corresponding directions (“go” and “stop”) (Numbers 1-2) Memorize two numbers & *switching* the corresponding directions (“stop” and “go”) (Numbers 1-2) *Can also change numbers “1 and 2” to “3” and 4”	Memorize one number & one corresponding direction (“stop”); All other numbers correspond to the opposite direction (“go”) (Numbers 1-5). Memorize 2-3 numbers & one corresponding direction (“stop”); All other numbers correspond to opposite direction (“go”) (Numbers 1-5).	Level 2 but switch the rules (Numbers 1-5) Level 2 but *two* numbers corresponding to (“stop”) (Numbers 1-10).
<u>Game 4: Clapping Game</u> Math: Counting, operations EF: Working memory, inhibitory control	Recall & reproduce counting (Numbers 1-5) Recall & reproduce counting (Numbers 1-10)	(Numbers 1-10) Recall & reproduce counting and addition with 1 Recall & reproduce counting and addition with 2 (Numbers 1-10)	Recall & reproduce counting but count to the number that comes one before (e.g. Subtract 1) (Numbers 1-10)
<u>Game 5: Number Share</u> Math: Numeral recognition, counting, one-to-one correspondence, cardinality EF: Cognitive flexibility	Counting out chips to that number (Numbers 1-5, or 1-10, depending)	Compare two number magnitudes after counting out chips (e.g. if numbers are green, which number has more? If numbers are orange, which number has fewer?) (Numbers 1-10)	Level 2, but add rule: “If numbers are purple…” determine how many counters to take away to make the quantities equal (Numbers 1-10)
<u>Game 6: Magic Cups</u> Math: Counting, one-to-one correspondence, operations EF: Working memory	Memorize number of objects (2-4 objects across two spatial locations)	Memorize number of objects (2-4 objects) when 1 is taken away or added to location (across 2 locations)	Level 2 but with 3 locations
<u>Game 7: Toy Search</u> Math: Counting, “counting on”, operations EF: Attentional control, cognitive flexibility	Search & “count on” objects that are similar across one trait (e.g. purple toys) (Up to 10)	Search & “count on” objects that are similar across two traits (e.g. purple *animal* toys)	Find sum of total objects “found” at the end by “counting on”
<u>Game 8: Match Pattern</u> Math: Patterning EF: Cognitive flexibility	Create AB pattern (by color; use same type of toys) Create ABB pattern (by color; can use different toy types)	Create ABB pattern (by type of toy, *ignoring* color)	Create ABBC pattern (by color) Create ABBC pattern (by type of toy, ignoring color)

<u>Game 9: Shape Search</u> Math: Shape recognition EF: Attentional control, cognitive flexibility	Identify one specific type of shape in a pool of other shapes (e.g. "circle all the triangles") Identify when two of that shape are next to each other (e.g. "circle when you see two triangles") (Shapes: Triangle, rectangle, circle)	Identify instances when one specific shape is followed by a different specific shape or vice versa (e.g. order of appearance does not matter; "circle when you see a triangle and rectangle next to each other") Shapes: Triangle, rectangle, circle	Identify instances when one specific shape comes first, and then the other specific shape (e.g. order matters; "circle when you see a triangle first, and then a rectangle") *Switch* the direction ("…now, circle when you see a rectangle first, then a triangle") Repeat Level 2 and above including: Trapezoid, rhombus, and finally, hexagon.
<u>Game 10: Shape Cards</u> Math: Shape recognition, shape attributes, magnitude comparison EF: Working memory, cognitive flexibility	Identify names of shapes & if that shape was the same as the previous shape they saw (e.g. n-back, 1-back)	Identify name of shape or number of sides of shapes, depending on the rule (the color of the shape denotes whether you identify the name or the number of sides).	Identify & compare number of sides of shapes with other shapes, depending on the shape color (one color: which has more sides? Other color: which has fewer sides?) Above, but switch color rule

Table III.2. Balance Check Comparing Treatment and Control Group Pre-Intervention

Matched sample (20 treatment, 20 control)	Treatment (M)	Control (M)	Raw difference (SE)	Standard difference
Male	.70	.60	.10(.15)	-.05
Household Income	50K	52K	2.00(.89)	-.01
Baseline EF	6.4	6.7	-.30(3.09)	-.03
Baseline math	9.8	9.8	-.00(.09)	-.00

F (5,34) = 3.28, *p*=.22

Table III.3. Child Response to Math/EF Activities

	Most popular (N children, out of 40 total children)	Least popular (N children)	Moved from Level 0-1, 1-2, or 2-3 during intervention (N children)	Did not master Level 1 at end of intervention (N children)	Mastered Level 3 at end of intervention (N children)
Game 1	17	3	33	7	13
Game 2	18	4	33	7	13
Game 3	39	0	33	2	15
Game 4	34	7	31	9	11
Game 5	2	14	29	11	8
Game 6	6	8	19	12	4
Game 7	8	0	37	0	21
Game 8	7	14	22	7	19
Game 9	13	15	20	14	11
Game 10	9	10	22	13	18

Table III.4. Treatment Impact on Math and EF

	Math			EF		
	Difference (SE)	*p*-value	Effect size	Difference (SE)	*p*-value	Effect size
Blocked sample	1.33 (.37)	.12	.06	9.73 (2.91)**	.003	.17

***p<.001, **p<.01, *p<.05, +p<.10

Table III.5. Time Spent in Different Intervention Formats Predicting Gains in Math and EF (Within Treatment Group)

	Math			EF		
	Unstandardized	β	*p*-value	Unstandardized	β	*p*-value
Pairs	.001(.01)	.01	.83	.05(.01)**	.32	.002
Centers	-.01(.01)	-.07	.29	.01(.04)	.10	.23
Circle time	.01(.01)	.07	.29	-.01(.03)	-.04	.73

***p<.001, **p<.01, *p<.05, +p<.10

CHAPTER IV: Feasibility of a Math/EF Intervention with Head Start Teachers and Coaches in Los Angeles

This chapter highlights the feasibility of coaching teachers on combined math/EF activities in the preschool classroom. The purpose of this chapter is to identify barriers that preschool teachers and coaches face with understanding math and EF development, the strengths and weaknesses of this approach in the classroom, and how to improve the coaching of these activities.

Theoretical Framework

Coaching is one avenue to support teachers' instruction and has gained momentum in early childhood education (Tout, Halle, Zaslow, & Starr, 2011). Coaching is generally defined as ongoing classroom modeling, observations, reflection, and evaluation of teachers (Sheridan et al., 2009; Whitebrooke & Bellm, 2013). The coaching model most commonly adopted by early childhood educators is program-level coaching, or the provision of training, classroom consultation, and other supports to strengthen the skills of administrators and teachers (Cohen & Kaufman, 2000). Within this model, researchers typically utilize a responsive coaching framework where coaches are encouraged to tailor their feedback to each individual teachers' instructional goals/needs and available classroom resources, and be actively involved in the modeling of practices and targeted teacher behaviors (Neuman & Cunningham, 2009). This framework is supported by theory on adult learning which suggests that adults engage in the learning process when the content and/or new knowledge relate to their current experiences in the learning process (Gordon, 2004). As such, many successful preschool coaching models have stressed the importance of a "two-way street" in the communication between teachers and coaches (Weiland et al., 2018). This might look like considering teachers' input on which approaches or curricular activities would work best given their classroom climate (e.g., would children be more engaged in large vs. small-group activities?) as this consideration has been shown to positively impact children's literacy gains from an intervention (Powell, Diamond, Burchinal & Koehler, 2010). Successful early education coaching models also tend to provide real-time data for teachers, as well as scripted instructional activities to better support teachers' implementation (Weiland et al., 2018).

Preschool Math Coaching: How Do We Measure Its Success?

In addition to adopting a responsive, program-level approach to coaching, it is also most effective if coaching is tied to a domain-specific curriculum (Piasta et al, 2017; Yoshikawa et al., 2015). To date, most domain-specific coaching efforts have been focused on literacy (Wasik, 2011; Powell, Diamond, Burchinal & Koehler, 2010; Wasik, Bond & Hindman, 2006; Hsieh, Hemmeter, McCollum, & Ostrosky, 2009) and fewer have focused on math (Clements & Sarama, 2011). This difference in attention is likely due to the lack of emphasis in philosophies of what matters in early childhood education. For example, most curricula focus on language, literacy and social-emotional learning rather than math (Weiland, 2016). Additionally, it is empirically clear that math learning matters for children's developing school readiness skills, even for domains outside of math, like reading (Purpura et al., 2017). This denotes a need in the literature to create and evaluate more coaching efforts targeting math teaching.

In Chapter III I already outlined a specific content model for targeting teachers' math instruction via integration with EF skills. The next step was then to identify the coaching targets and measures to quantify the success of coaching. The most common method of determining the effectiveness of early childhood coaching is by collecting observations of teachers' fidelity of implementation or changes in teaching practices, engagement/enthusiasm of teachers (Pianta et al., 2014) and impacts on child outcomes (Hamre et al., 2010). However, in addition to these three indicators of coaching success, Guskey (1994) theorized one additional categories: documented gains in acquired knowledge of the participants (e.g., teachers). The aforementioned fidelity of implementation and child impact indicators provides researchers with only an external indication of coaching effects, but does not provide any indication of whether a teacher deeply understands the purpose of the practice and has acquired meaningful knowledge beyond being able to (at its most basic level) mimic a practice that has been modeled for them. Evidence for gains in teacher knowledge and understanding of a curriculum—a more *internal* indicator of coaching success—are not typically studied in evaluations of coaching models in early childhood education. Although teacher understanding and teacher practice are theoretically related, we still have little evidence for the extent to which they are empirically related in the context of coaching.

To date, measures of teacher knowledge in preschool are typically used to explain differences in teachers' instructional practice and students' academic achievement, independent from coaching teachers on implementing a new curriculum (Cunningham, Zibulsky, & Callahan; Lee, 2017; McCray & Chen, 2012; Piasta, Connor, Fishman, & Morrison, 2006). There are some examples in the preschool literature, however, that have investigated how professional development and coaching have impacted teacher knowledge—in addition to fidelity of implementation and child outcomes. For example, one evaluation of a professional development program that introduced a science curriculum in preschools found that teachers improved in their science teaching knowledge and demonstrated higher quality science teaching compared to control classrooms (Gropen, Kook, Hoisington, & Clark-Chiarelli, 2017). For preschool math coaching, however, Clements and Sarama (2011) have included “discussions” within their professional development series (TRIAD model) to provide an opportunity for teachers to ask questions about math content and knowledge (e.g., what skill comes next in a child’s development?), but there was no concrete assessment of teacher knowledge throughout the course of the intervention. Taken together, it would be useful to examine how math coaching influences teachers’ knowledge and their fidelity of implementation within the same study.

Coaching in the Los Angeles Head Start Context

A math-specific coaching model that gathers data on teacher knowledge could be especially beneficial for the Head Start Los Angeles context. In a random subsample of Los Angeles County public preschool teachers (*N*=59), nearly half of them reported that math and other STEM activities were not a priority in their classroom and that they had limited access to STEM instructional materials and training (Barrett, 2017). A regression discontinuity design revealed that children who attended Los Angeles County preschools did not make statistically significant gains in math compared to children who just missed the age cutoff for preschool enrollment (Kyger & Barnhart, 2017).

However, Head Start teachers across the United States have reported being overwhelmed with implementing instruction (including math) because of the competing task of addressing children’s social, emotional, and behavioral needs (Jacoby & Lesaux, 2017). In Los Angeles specifically, math instruction may also be competing with Head Start teachers’ focus on supporting children’s English language learning (81% of students; LACOE Annual Report, 2019). Furthermore, Head Start centers in Los Angeles only receive coaching support to improve teachers' CLASS scores as part of the mandatory Quality Rating of Improvement (QRIS) assessment requirement for Head Start. Yet coaching on the CLASS only provides general, non-content specific support to teachers. Past evaluations of coaching to improve CLASS scores has demonstrated null impacts on direct assessments of children's academic achievement (Pianta et al., 2017). Due to this, Los Angeles Head Start teachers may be less likely to adopt evidenced-based math practices because of the behavior and language needs of their students, but a responsive coaching approach that tackles math and children's self-regulation (and underlying EF skills) may be promising for this population. Additionally, collecting data on teacher understanding of a math/EF approach and identifying barriers to implementation could help inform the scalability of this effort moving forward. In the current study, I examined the feasibility of a revised version of the math/EF activities from the previous chapter with Head Start teachers and coaches in Los Angeles.

Child 360 partnership. For this study, I partnered with Child 360, a nonprofit organization based in Los Angeles that strives to "advance early education program quality and capacity by supporting the development of the whole child, growing a qualified and diverse workforce, and strengthening family engagement." Child 360 provides coaching support for teachers, quality assessments, family engagement support, and workforce and professional development trainings. In the 2015-16 year, Child 360 served 256 preK programs (78.5% center-based; 21.5% family child care) and 9,603 children. Over 50% of the children in Child 360's network have a household income of $29,000 or less and 63% of children are dual language learners. Child 360 offers instructional coaching if partnering programs elect to have this service free of cost. Child 360 is funded by First 5 LA and the California State Preschool Program QRIS Block Grant. Child 360 has 30 instructional coaches with a coach-teacher ratio of 1:26. The

Child 360 Research & Evaluation team is specifically in charge of collecting data on classroom quality assessments (CLASS) and collecting child outcomes on literacy (Preschool Early Literacy Indicators; PELI) and math (Individual Growth and Development Indicators of Early Numeracy). They also recently developed a "STEM Initiative" aimed at improving the amount and quality of STEM related instruction in the classroom and opportunities in the home. I am partnering with Child 360 to target the "M" (math) of "STEM" and to provide more instructional support to teachers in this domain.

Research Aims

1.) How does the math/EF coaching improve teachers' and coaches' understanding of preschoolers' EF skills and math development?

2.) What is the overall quality of the coaching model (e.g., teacher ratings of coach quality, teacher reported knowledge gained from coaching sessions)?

3.) What are the overall strengths and weaknesses of the content and logistics of this intervention from the perspective of the coaches and teachers? (e.g., feedback on activities, feedback on frequency and method of coaching)

4.) What is the *overall feasibility and promise* of this coaching model (e.g., teacher interest, teacher compliance, teacher confidence, teacher fidelity of implementation, teacher report of child learning gains)?

5.) Post-hoc: How does teachers' knowledge and understanding of EF integration and math skill development relate to their fidelity of implementation, education, and years of teaching experience?

Method

Recruitment

Child 360 invited six Head Start centers to participate in this intervention, with the hope that at least five would agree. Fortunately, all six Head Start centers (and 10 teachers) agreed to participate. Each center had an existing Child 360 instructional coach who had been helping the teachers within that center on improving their instructional support, as part of improving their overall CLASS scores (*N*= 4 coaches). Thus, the coaches already had established relationships with the teachers. For this study, Head Start centers were told they would be receiving instructional coaching to support their teachers' math instruction. Teachers and coaches were each compensated $400 for completing surveys and participating in an interview. Coaches additionally received reimbursement for gas mileage.

Teacher Demographics

All ten teachers were female and ranged in age from 27 to 63 years (*M*=42 years). Nine teachers identified as Hispanic and one teacher identified as Black. There was a wide range of teaching experience, with one teacher in her second year of teaching preschool and another teacher in her 23rd year (*M*= 5 years). Three teachers had a two-year (Associates) degree, five teachers had a four-year (Bachelor's) degree, and two teachers had a Master's degree.

Head Start Centers

All six Head Start centers were representative of the larger Head Start population in Los Angeles county and were located in four major regions of Los Angeles. One center was from Exposition Park (South LA), two centers were from Downey (East LA), two centers were from North Hollywood (North LA), and one center was from Inglewood (West LA). Seven Head Start classrooms (four centers) were full-day programs and three classrooms (two centers) were half-day programs. The structure of the full day programs included breakfast (7:45am), brushing teeth, 10-minute large group, 10-minute small group, 50-minute center time, 50-minute outdoor time, lunch, nap, read-aloud, and more outside time. The structure of the half-day programs included breakfast or lunch, brushing teeth, 10-minute large group, 10-minute small group, 50-minute center time, and 50-minute outdoor time. For the two centers that implemented half day programs, student days were Monday through Thursday and Fridays were teacher-only days. On average, each Head Start classroom had *N*=15 children (range of 9-15 children were present each day of in-classroom coaching; *M*=13).

The Math/EF Activities (Revised from Study II)

The Math/EF activities were improved after results from my pilot study (Study II). One of the critiques in my presentation of the pilot study activities (from colleagues at the University of Michigan in 2017 and my presentation at the Pacific Coast Research Conference in 2018) was that the activities were more EF-focused than math-focused. To address this, I refined my activities to include more math topics and levels. An overview of the activities is in Table IV.1.

Number sense. I created a total of five activities that developed children's number sense skills. In the Number Search activity, children practiced basic number recognition. A series of laminated number cards were spread out in front of a child and they were asked to search for a specific number (e.g., "3") and teachers could make it more difficult by asking the child to look for all the "green 3's", all the "small 3's", or all the "small green 3's". This exercised children's *cognitive flexibility* (switching between what they are searching for), *working memory* (remembering what they are supposed to be searching for), and *response inhibition* (inhibit the response to pick up all the 3's when teacher says to pick up only the green ones). Number Music was also a number recognition game where there were numbers in a circle (Velcro numbers on the carpet) that children would walk around while music played. Then, when the music stopped, children had to freeze, and when the music started again children had to pick up their number from the carpet and try to match it with a peer's number. This activity exercised children's *response inhibition* (freeze when music stops) and *working memory* (remember the task). Dot Fun was an one-to-one correspondence, cardinality, and magnitude comparison activity where children had to count how many big/small, green/blue/red dots there were on a card, as well as compare which group of dots had more/fewer. This activity exercised children's *cognitive flexibility* (switching between the rule) and *response inhibition* (inhibit counting all the dots and just focus on a specific color). Comparing Blocks was also a counting, cardinality and magnitude comparison game where peers had to compare number of toys (or drastically different sizes-big/small) to see which person had more or fewer. This exercised children's *response inhibition* (objects were of different sizes so child had to ignore the size and focus on the quantity). Lastly, the Freeze Game helped children with counting and counting on where children had to crouch upwards with their bodies and freeze at the number the teacher said to freeze at (e.g., 4), and the game got more difficult as children had to not only freeze at a certain number but count-on from that number to another number (e.g., starting at 4, count to 6). This exercised children's *response inhibition* (have to inhibit response to keep counting all the way to 10) and *working memory* (have to keep track of what number to freeze at).

Operations. Monkeys in the House focused on addition and subtraction where children had to keep track of how many monkey counters (or any other animal counter) entered a wooden block house (added one at a time; up to 4 monkeys), and then the teacher would add 2 at a time, subtract 1, and then increase the difficulty by adding two houses. This exercised children's *working memory* (have to keep track of how many monkeys in the house at any given time). The Comparing Blocks game also had an operations component where children had to decide how many more toys their peer needed so that they had the same number of toys. For these activities that stressed operation skills, teachers used a ten frame with magnetic dots to scaffold children's learning of "one more" and "one less."

Geometry. Shape Search was an activity that is similar to Number Search where children had to search for a given shape based on its name, size, color, or number of sides. This activity exercised children's *cognitive flexibility* (switching what they were searching for). And the Shape Sort built off of this game where children had to not only search for shapes, but then sort shapes based on different attributes (e.g., color, size, number of sides etc.). This exercised children's *cognitive flexibility* (switching what rule they were sorting on). Constructing Shapes focused on children's knowledge of closed sides and corners that made a shape where children built a triangle with three short sides, three long sides, two long/one short, one long/two short, and then had to move on to building rectangles with two long/two short, or a square with all short or all long sides. This exercised children's *cognitive flexibility* (switch between what shape they were building and if they were using short or long sides) and *working memory* (keep in mind what shape they were building and sides to use). Lastly, Shape Pattern Blocks used the popular shape pattern blocks or Tangrams (and accompanying templates) for children to practice fitting shapes together to form a picture. The EF component occurred when teachers first had children use colored shape blocks that matched the colored shapes picture, and then they swapped out these shape blocks for colors that did not match the shapes in the picture. Thus, children had to practice *attentional control* (ignore the color and focus on the shape), and teachers also asked children to count how many blocks of a certain shape they used which exercised *cognitive flexibility* (had to switch and focus on a specific shape when counting).

Measurement. The Measuring Worms activity involved the teacher first reading aloud a book about an inchworm measuring all his animal friends, and the teacher demonstrated measurement by using an inchworm counter to measure each animal in the book. Then, children practiced measuring each other, where one child laid down in the middle of the carpet and the teacher used 4-inch worm counters to measure how long the child was, and then had another child lie down next to the line of worms to see if they were “longer” or “shorter” than the first child who was measured (and if more worms needed to be added or subtracted). Then, children went to small group to practice measuring laminated animals of different lengths, ranging from two inches to eight inches (bunny, turtle, elephant, mouse, tiger, bird). At first the child used the 1-inch worm counters, and then they measured the same animal using the 2-inch, 3-inch and 4-inch counters. This exercised children’s *cognitive flexibility* as they had to engage in trial and error with the different units of measurement to fit the length of the animal. They also had to be able to articulate that two animals could be the same length even though one was measured using three of the “baby” 1-inch worms, and another animal was measured using only one of the 3-inch worms. This activity was practiced over a series of days so that teaches could carry out the developmental progression of the activity by first introducing children to measurement, and then helping them understand different units, and finally demonstrating that it takes more of the “baby” worms to measure the same object because the baby worms are shorter.

Patterning. For patterning, originally, the activity was patterning with animal and fruit counters and switching between whether they were patterning by color or animal/fruit (Study II). However, this activity was changed after meetings with the Child 360 instructional coaches who noted that teachers generally lack understanding of children’s patterning skills and using the “AB” concept when teaching patterns. Thus, for this activity I created different patterning templates of different colors. For the first level I had a template with squares of two different colors labeled “A” and “B” where children would just copy the template with their own Unifix cubes. Then, children copied the same patterns but with "AAB" or "ABB" templates, and then children used black and white AB and ABB/AAB templates, and finally they used templates without any colors. Once children mastered the latter level the teacher would encourage them to use a different color for "A" and "B". This exercised children's *cognitive flexibility*. Then, children were given different colored templates that did not match the "A" and "B" letters on the template that children had to follow. This exercised children's *response inhibition* and *attention control* (children had to ignore the color and focus on the AB pattern they were supposed to make). Lastly, children were encouraged to learn and practice ABC patterns, create their own patterns, and teachers continued to supply them with the "EF" color/letter mis-match templates.

Procedure

An overview of the study design is in Figure 1. Perfect fidelity to the study design was almost reached. However, there were a few instances that deviated from the intended procedure that are noted throughout this section, namely with (1) a delay in timing of certain study activities and (2) infrequency of visits from instructional coaches.

In September 2019, the four coaches from each center met with me to receive a two-day training (4 hours total) on the EF and math activity booklet. I designed this coach training with one of the coach supervisors at Child 360. On the first day, I provided an overview of executive functioning and how this skill was exercised in young children and then highlighted the different math topics covered in the booklet and the developmental levels within each activity. At the end of the first day, I modeled all of the activities for the coaches. On the second day, the coaches practiced these activities on their own with my oversight where one coach played the role of the teacher and the rest participated as students. The coaches also completed their first survey at this time.

In the first week of October 2019, I jointly observed with a classroom's assigned coach each of the classrooms for one full day to assess math instruction and the availability of math materials. In the last two weeks of October, the assigned classroom coach and I met with each teacher to introduce the math/EF activity booklet. At this meeting, the teachers also completed their first survey (pre-coaching survey). We also asked their input on how the activities could fit into their school day and which materials could be adapted for these activities. I assisted teachers in gathering the materials together and provided them with the additional materials that were needed. At this meeting, I also rated teachers' attitude and interest in the coaching on a scale of 1-3 (not interested, mildly interested, very interested). Perfect fidelity was not reached in this part of the study because I would have liked these activities to happen at the end of September/early October, but due to scheduling, the observations and coach/teacher meeting lasted throughout the month of October.

In November 2019, I visited each teacher 2x per week to model activities for the teacher during a 15-minute large-group block. The coach visited once every two weeks to model an activity. Each visit throughout the duration of this intervention averaged around 33 minutes (range 17- 48 minutes). In November and December 2019, a total of eight activities were modeled for teachers in the classroom with children. In terms of pacing, perfect fidelity was not reached because the first four activities took the longest time to model for the teachers. This was in part because the teachers were not used to in-classroom modeling and instead received feedback after an in-classroom observation. More than half (*N*=6) of the teachers perceived the math coaching as more of an enrichment experience to benefit the children, not to necessarily teach them (the teachers) to improve upon their math instruction. For example, when I modeled the activities during large-group, the lead teacher would often go about organizing something else in the classroom. This miscommunication was sorted out after the first few modeling visits, but with some teachers it still remained an issue, so the coach and I had to model the first four activities more times than the other activities. The modeling of the first four activities also took longer because both the coach and I modeled *each* of them for the teacher. After this, however, we found it more efficient to "split" the modeling of activities. Ideally, I was hoping for the coach to model all of the activities for the teacher and I would just jump in when needed. But I soon learned that the coaches were realistically only able to visit about once every two weeks, which did not work for my timeline. Thus, at the end of the modeling period, I ended up modeling 7-8 of the activities for the teachers, and the coach only modeled 5-6 of the activities. Right before the holiday break (December 14-18, 2019) and in some cases after the holiday break (January 6-7, 2020), I met with the teachers individually to complete the mid-coaching survey. Perfect fidelity was not achieved in this part of the study because, again, activities were delayed due to prior delays, and I was not able to model activities until November (not October). Additionally, coaches did not visit teachers 1x per week (as I planned), and instead visited 1x every other week.

In January 2020, the coaches continued to visit teachers 2x per month, and I reduced my visits to 1x per week. Teachers were expected to implement two math activities per week (for one of these activities I observed and modeled if needed), and teachers filled out weekly activity logs to report on the activities they did that week. In February 2020, I introduced the last four activities that were all implemented at small-group (patterning, constructing shapes, composing shapes, and a measurement activity). Teachers then practiced and implemented these activities for the month of February. Provided that we were delayed from the start of the study and coaches continued to visit once every other week, I was nevertheless pleased with how much progress teachers made with the learning and implementation of the activities.

In March 2020, I collected fidelity of implementation video footage of each teacher. The plan was to collect five activity implementation observations per teacher but because of the COVID-19 related school closures in mid-March, I was only able to record one observation for seven teachers and two observations for two teachers (total *N*=9 teachers had at least one observation). Thus, perfect fidelity to this portion of the study was not achieved. Additionally, in March and April 2020, I was not able to conduct a final classroom observation, but I was able to virtually interview teachers and coaches (individually) and emailed them the final survey.

Measures

Classroom observation of math instruction. I created a math observation measure for the purposes of this study to assess teachers' baseline math instruction and math opportunities provided in the classroom (see Appendix B.I). I also planned on administering this observation measure at follow-up, but due to COVID-19, I was not able to conduct a follow-up observation. This measure documented the observed math instruction throughout the school day, such as whether there was a math center, a math small-group activity, and a math large-group activity (yes/no), and the nature of these activities (e.g., the specific math skill(s) targeted). I also recorded whether or not there was a teacher present (in the case of the math center), number of students at each table (in the case of small-group), number of children that participated (in the case of center-time), and whether the teacher provided any kind of instructional support or direction to children at center time (yes/no). I also took notes on the available math materials and whether there was any teacher math instruction or "talk" during meal-time and transitions. I also took notes on whether there were any observed instances where instruction was too difficult, poorly executed, or vague (no child was able to respond to the question e.g., "what is 9 boy plus 6 girls?"; or "How come this shape does not fit into this shape?").

Variables created. Since I was unable to conduct a follow-up observation, I did not end up using the data from this measure to answer any of my research questions (e.g., whether teachers improved in their overall math instructional practice pre to post). Instead, I used data from this measure to paint a descriptive picture of the *variability* in the sample of teachers on their math instructional practices. To compare teachers, I created variables that were common across all classroom observations (e.g., small-group and large-group math were excluded because not all classrooms had math activities scheduled on that day). For example, all classrooms were able to be compared on (1) the availability of math materials (1= lacking, 2= adequate, 3= in abundance), (2) the quality of the math center (1=center is "open", but no materials were out on the table, 2= materials were out, but no teacher was present/no instructional guidance on how to interact with materials, 3= materials were out and teacher was present to provide instructional guidance), and (3) how many teachers committed instructional errors in the following identified categories: practicing numeral recognition above 10 before children established cardinality, children copying patterns without teacher highlighting the repeating unit, addition/subtraction with sums greater than 10, using shape blocks without practicing shape names, using open-ended questions that were too difficult/children were not able to answer).

Coach surveys. The coaches completed one survey (see Appendix B.II). This survey was administered after I introduced the math/EF activity booklet and modeled the activities for the coaches. This survey collected information on how well the coaches understood the activities' math levels and EF integration, their comfort-level with modeling these activities for the teacher, and any obstacles they anticipated encountering during coaching. The survey took no longer than 10 minutes to complete.

Variables created. From this survey, I created one variable for coaches' understanding of the math levels (rating scale 1-4) and one variable on coaches' understanding of the EF integration (rating scale 1-4). My rating rubric for quantifying coaches' understanding of math and EF is in Appendix B.XI.

Teacher surveys. Teachers were surveyed pre-coaching, mid-coaching, and post-coaching. The pre-coaching survey gathered demographic information (such as age, race, years of experience teaching preschool, and their level of education), teachers' past experience with math coaching, teachers' overall confidence in teaching math, and the curriculum or resources teachers used for math activities (see Appendix B.III). The mid-coaching survey was administered after teachers had received modeling but before implementing the activities on their own (see Appendix B.IV). This survey collected data on teachers' feedback on the modeling of the activities, their understanding of the math levels and EF integration of four activities, and their confidence in being able to implement the activities on their own (up until this point they had only received coach modeling). The post-coaching survey collected the same information as the mid-coaching survey (see Appendix B.V), except understanding of EF integration and math levels was assessed on all of the activities, and not just the four they were assessed on mid-coaching. The pre- and mid-coaching surveys took around 20 minutes to complete and the post-coaching survey took about 45 minutes.

Variables created. Similar to the coach survey, for the mid- and post-coaching surveys, I created one variable for teachers' understanding of the math levels (rating scale 1-4) and one variable on teachers' understanding of the EF integration (rating scale 1-4). I also created two variables on math teaching confidence from the pre-coaching and post-coaching survey (rating scale 1-3). From the post-coaching survey, I also created a variable on how teachers rated their coaching experience on all 12 math/EF activities (rating scale 1-4). The post-coaching survey also gathered descriptive information on how often teachers preferred to be coached (e.g., weekly, bi-weekly etc.), as well as how they liked to be coached (e.g., modeling-observation-feedback, just modeling- no observation, in-classroom vs. out of classroom coaching etc.). At the time of the pre-coaching and post-coaching survey, I also created an attitude/interest variable that captured teachers' interest in the math coaching on a scale of 1-3 (1= not interested, 2=mildly interested, 3=very interested).

Teacher activity logs. Once teachers were implementing activities entirely on their own (January 2020), they also completed weekly activity logs for 10 weeks that documented which activity they implemented, ratings of child engagement/participation, the overall success of the activity, any evidence for specific child learning gains, and noted any struggles in implementing the activities/questions for the coach. The activity log took no longer than 10 minutes to complete (see Appendix B.VI).

Variables created. The purpose of activity logs was for reference at coach meetings and to help teachers keep track of the activities they administered on their own. The only quantitative variable created from these activity logs was teacher log compliance (1=teacher completed fewer than half the logs, 2=teacher completed more than half of the logs but not all of them, and 3=teaches completed all of the logs). I also used teacher activity logs for qualitative evidence of children's learning gains. For example, teachers made note of the specific children they worked with that day and what level the child(ren) started on and ended on at the conclusion of the activity. Teachers were then encouraged to start on this level the next time the child did that activity and make notes of any child learning gains they observed.

Fidelity of implementation. This measure assessed how well the teachers were administering the activities (see Appendix B.VII). Each coach was able to videotape one teacher (*N*=4 teachers), and I videotaped a total of 9 teachers. Video footage was then rated on organization, adherence to script, quality of math skill level, quality of EF integration, and child participation for each activity (For each dimension, teachers were giving a rating of (1-3)). Two coaches and I rated each video (*N*=3 scores per teacher). I trained coaches on the rating system by demonstrating the system on one observation video (a pilot fidelity observation I conducted in December for one teacher- not used in analyses), but other than that, there was no reliability established prior to coaches rating videos. Instead, I calculated reliability coefficients for each domain after ratings were complete. On average, reliability was good for scores on child participation, EF integration, and organization (kappa's > .90), reliability was acceptable for adherence to the script (kappa = .74) and reliability was poor for identifying appropriate math skill level (kappa = .48). Because of this, I averaged the scores (my score and the two coaches' scores) for the reliable components of the measure and only used my score for the unreliable component of the fidelity measure (math skill level). It is likely that coaches were not reliable on their rating of appropriate math skill level because as later evidenced, coaches did not score highly on their knowledge of math skill(s) development. See Appendix C.VIII for the rating rubric used to assess fidelity of implementation.

Variables created. For fidelity of implementation, I created five variables for each component of fidelity on a rating scale of 1-3: organization, adherence to script, quality of math skill level, quality of EF integration, and child participation in each activity

Semi-structured interviews. This measure was developed for the purposes of this study and consisted of a list of questions that probed teachers' and coaches' perceptions and insight into the intervention and coaching model. I focused on (1) how this intervention and coaching changed their understanding of math and EF development; and (2) the strengths and weaknesses of the intervention content and coaching model. Although there was a pre-set list of questions, since this interview was semi-structured, the interviewer (me) had the liberty to ask clarifying or follow-up questions after a teachers'/coaches' response. Each of these interviews was conducted individually and virtually. These interviews occurred at the conclusion of intervention/coaching (range March 30 – April 23, 2020) and each interview lasted around 45 minutes (see Appendix B.IX and B.X). Detail on the data analytic approach in creating variables from these interviews is provided in the next section.

Data Analysis Plan

Aim 1: Teacher and coach understanding of EF/math content. To address the strengths and weaknesses of the content of the intervention, I examined the means and standard deviations of teachers' and coaches' understanding of the content of the intervention—specifically with their understanding of the developmental levels within each math activity and how EF was exercised within each activity (rating scale 1-4). I also calculated changes in the mean on these two variables from the mid-coaching time point to the post-coaching time point.

Aim 2: Teacher response to coaching model. To address the promise of the coaching model, I calculated means of teacher-reported quality of coaching received per activity (1-4), and the percentage of teachers who requested more coaching on each activity. I also qualitatively analyzed emerging themes of teacher's gains in knowledge of math and EF development (from the semi-structured interview; analytic details presented in Aim 3).

Aim 3: Strengths and weaknesses of the content and coaching. To address the overall strengths and weaknesses of the coaching and content of the intervention from the perspectives of the teachers and coaches, I analyzed interview data. I employed a qualitative approach using The Framework Approach (Pope, Ziebland, & Mays, 2000) that charts themes and categories for "positive" and "negative" aspects of the 1) content of the intervention and 2) coaching model. This approach has been previously used in studies analyzing the efficacy of preschool curricula with teachers (e.g., PATHS curriculum; Hughes & Cline, 2014) to identify "positive" and "negative" components of the curricula. This Framework Approach lies within a broad family of analysis methods termed *thematic analysis* or *qualitative content analysis*, which aims to draw conclusions from data clustered around themes or categories.

To code these data, I transcribed the interview and two undergraduate RA's identified phrases in the data that were mentioned more than once by teachers and I categorized these phrases into themes. This process was used in analyses of the Head Start Cares' participant interviews (Mattera, Lloyd, Fishman, & Bangser, 2013).

Aim 4. Overall promise and feasibility of intervention. To analyze the overall promise and feasibility of the intervention, I was calculated descriptive statistics (means) of teacher attitude/interest (rating 1-3) and the change in this attitude pre to post intervention, teacher confidence in teaching math (rating 1-3), and change in this confidence pre to post intervention, teacher activity log compliance (rating 1-3), child engagement (rating 1-3), and overall fidelity of implementation (rating 1-3). I also analyzed qualitative data from the teacher and coach interviews on critiques of the intervention and coaching, and evidence for child math learning gains (from the teacher activity logs).

Results

Baseline Teacher Math Practices

Baseline observation data on existing math instructional practices of the 10 classrooms is in Appendix C.

Teacher reported math practices. Four teachers reported that they used math activities from Creative Curriculum, and six teachers reported using a mixture of Creative Curriculum and their own activities adapted from resources found on the internet or from another teacher. Four teachers reported no prior coaching on math teaching, and the remaining six teachers reported that they had only received coaching on how to increase math-related language in the classroom without suggestions for specific instructional activities. Seven teachers reported that they implemented a small-group math activity once per week, and three teachers noted that they usually did a small-group math activity every other week. For large-group math activities (beyond counting on the calendar or taking attendance), four teachers reported that they had done number or shape recognition activities, whereas the remaining six teachers reported no instructional math activity at large group. On a scale of 1-4 in confidence in teaching math (not confident, somewhat confident, confident, very confident), teachers scored an average of 2.7 (range 1-4).

Aim 1: Teacher and Coach Understanding of EF/Math Content

Coach understanding pre-coaching. Results from the coach survey revealed that coaches were generally able to articulate how EF was exercised in four of the activities that emphasized different EF components (Freeze Game which stressed inhibitory control; Monkeys in the House which emphasized working memory; Number Search which emphasized attentional control; Shape Sort which emphasized cognitive flexibility). For example, in the survey I rated each of the four EF activity responses on a scale of 1-4 (rubric is in Appendix B.XI), and on average, coaches scored a three (3= "Was able to connect the specific "EF" aspect of the activity that and used correct terminology when describing this connection, but the description was vague) out of four. In terms articulating the math skill sequence, for four activities that covered a range of math skills (patterning, geometry, operations, number sense), coaches struggled more with this (*M*=1.75, out of 4) ("1= Can identify only one level correctly"; "2=Can identify two levels").

Teacher understanding mid-coaching. After receiving modeling of the activities, the teachers were also tested on their understanding of how executive functioning was integrated in each of four activities and their understanding of the math levels on four different math activities (identical to Coach survey). Results revealed that unlike the coaches, teachers did not score well on articulating how executive functioning was exercised (*M*=1.4) (in between scores "1"=May use EF-related language but does not connect this to any specific facet of the activity e.g., "Children have to pay attention to the rules" and "2"=Connects EF-related language to a specific facet of the activity- but it is the wrong facet; Or identifies the correct facet but describes the type of EF incorrectly). However, similar to the coaches, teachers also did not score well on identifying the different math skill levels of each activity (scale of 1-4; *M*=1.9).

Teacher understanding post-coaching. A summary of teachers' understanding of math development and EF integration across four selected activities is detailed in Table IV.5. Compared to the mid-coaching survey, only four teachers improved slightly in their explanation of how EF was integrated in four activities (*M*=1.9 compared to *M*=1.4), and seven teachers improved in their articulation of the sequence of math skills in each level of four selected activities (*M*=2.6 compared to *M*=1.9). Post coaching, teachers were best at articulating the EF integration of the first three activities they learned (scored 3 < out of 4; Number Search, Freeze Game, Monkey House). Post coaching, teachers were also best at articulating the developmental levels of the shape and patterning activities (scored 3 < out of 4; Shape Search, Shape Sort, Patterning), but were less able to articulate developmental levels in the number sense and operations topics.

Aim 2: Teacher Response to Coaching Model

Ratings of coaching quality. Also referenced in Table IV.5 are the teacher ratings of the quality of coaching they received for each activity and if they wanted to receive more coaching on a given activity. On average, teachers rated the coaching quality across activities a 3.6 out of 4. The highest rated coaching was recorded for the Number Search, Freeze Game, Shape Sort and Patterning activities (average 4/4). And across all the activities, on average, 8 out of 10 teachers were satisfied with the coaching of that activity and did not express needing more clarification or coaching on the activity.

Teacher reported gains in knowledge. During the interview, teachers also reported on what they learned from the coaching. Two themes emerged that all ten teachers commented on: (1) how the coaching increased their math vocabulary and (2) that they gained knowledge of children's math learning readiness. In Table IV.8 are example quotes from teachers of what they learned from this math coaching. Specifically, for math vocabulary, five teachers commented on learning terms such as "cardinality," "one-to-one correspondence," and shape names like "trapezoid" and "rhombus." Two teachers commented on learning new terms of measurement (e.g., "long/short" instead of "big/small"). For math learning readiness, teachers reported learning that children are ready for other geometry topics beyond naming shapes (N=3), that children's patterning skills are ready to be expanded beyond simple copying (N=4), and that children are ready to learn addition and subtraction (*N*=2). Six teachers also noted how they did not know that most children in their classroom were not able to work with quantities above five despite being able to verbally count above five (e.g., counting and cardinality are not necessarily the same thing). Four teachers commented on how they learned that "how many?" is a key question to ask when assessing whether a child has understanding of cardinality and not just counting ability. One teacher also commented on how measurement can be taught in a different way beyond just conventional units (e.g., using a tape measurer), and another teacher expressed how impressed she was that children had such good working memory at such a young age.

Aim 3: Strengths and Weaknesses of the Content and Coaching

A summary of strengths and weaknesses of the content and coaching from the teachers' and coaches' perspective are summarized in Table IV.6.

Teachers' response to coaching. From the teacher interview, themes emerged regarding the strengths and weaknesses of the math coaching. For strengths, one theme emerged: overall, teachers liked the in-person in-classroom modeling, and all ten teachers expressed no interest in virtual coaching, out-of-classroom modeling without in-classroom modeling, nor only being provided with materials and the activity booklet without coaching. Five out of the ten teachers also indicated that they liked the in-person modeling followed by observation and feedback of their practice. The remaining five teachers, on the other hand, expressed that they preferred to only have the in-person in-classroom modeling and not have follow-up observation and feedback (because the observation itself was stressful for them).

For weaknesses of the coaching, three themes emerged: lack of coaching for how to support children of the lowest skill levels, disruption of teaching flow, and coaching frequency. For the first theme, seven teachers expressed that they understood how to progress through the levels of the math activities, but often times some of the children would get stuck, and not even comprehend the first level very well. In future math coaching, teachers expressed wanting more coaching on how to help these lowest-skilled learners. For the second theme—disruption of teaching flow—six teachers commented on how it was difficult to manage children's behavior in the classroom while simultaneously being coached, five teachers commented on how the coaching sometimes disrupted their lesson plan for the week, and five teachers recommended having more modeling sessions outside of the classroom or when children were not in the classroom. For frequency, six teachers did not like the weekly coaching and preferred either every other week (3 teachers) or monthly (3 teachers).

Teachers' response to content. Table IV.7 documents specific quotes from teachers on strengths and weaknesses of each activity. However, some themes emerged across the entire intervention content (all 12 activities) more generally. In terms of strengths, at least six teachers commented on how the variety of materials and math topics was helpful, children were very engaged in most activities, and breaking down a math activity into different skill levels was helpful. Teachers' suggestions for improvement on content included having more materials for small group (at least six teachers noted this). For example, for the Constructing Shapes activity, there were enough magnetic sides and angles for each child to make a triangle, but when they had to make a rectangle, they had to borrow and take turns with their peers. Teachers also suggested creating more introduction activities to help introduce foundational concepts—specifically for geometry and measurement. For example, many children could not identify any shapes, which made it difficult for them to search for and sort shapes.

Coaches' response to coaching and content. From the coaches' perspective, the detailed activity booklet and accompanying scripts for each of the activities was the one emerging theme for strengths of the content of the intervention. For weaknesses, the coaches highlighted that the activities would benefit from being explicitly tied to goals of the Classroom Assessment Scoring System (CLASS), the Early Childhood Environment Rating Scale (ECERS) and the Desired Results Developmental Profile (DRDP), considering coaches already utilized these tools. A strength of the coaching from the coaches' perspectives is the actual coach training itself, and having them participate in the modeling and education of the materials. Overall, coaches commented on how it was rare for them to actually model full-fledged activities in their typical instructional coaching, and that acquiring a deep knowledge of executive functioning and the math skills was really useful.

Aim 4. Overall Promise and Feasibility of Intervention.

In order to assess the feasibility of this intervention with teachers I examined teachers' implementation of the activities on their own (fidelity of implementation), their attitude and interest in the coaching, their confidence in administering these activities moving forward, and qualitative evidence of child learning gains.

Fidelity of implementation. A summary of teachers' fidelity of implementation scores is in Table IV.4. For adherence, four teachers followed the script exactly, three teachers followed the script generally but also added in their own elements, and two teachers strayed from the script substantially. For EF integration, five teachers implemented it fully, three teachers implemented it partially, and only one teacher did not implement the EF integration. For math skill level, two teachers started at the correct level and moved on to the next level when most children were ready, four teachers started at the correct level but did not move on quickly enough or moved on too quickly, and two teachers started at a level that was too difficult and had to restart and begin with an easier question for some children. For organization, four teachers were organized with materials and most children were on-task, four teachers were either disorganized with materials or only focusing on one or two children during the activity—leaving other children to become bored or off-task, and two teachers did not have materials set up and children were off-task for a majority of the activity. For child engagement, once materials were organized, seven of the teachers had at least half or more than half of the children engaged whereas the other two teachers had more than half of the children not-engaged.

Teacher attitude/interest. During my first meeting with the teachers, I observed their questions and intrigue in learning, and their overall demeanor towards my being there. Three teachers were 1=*not interested/annoyed* in the math coaching, four teachers were 2=*mildly interested/not annoyed but not enthusiastic either*, and three teachers were 3=*very interested/enthusiastic*. At the conclusion of the study, only one teacher was not interested/annoyed, three teachers were mildly interested, and six teachers were very interested/enthusiastic. This represented an increase in interest/attitude of .40 (M=2.1 to 2.5) of the teachers. Teachers also reported on their enthusiasm towards the activities by rating their top three favorite activities (Table IV.5). There was no clear "winner" for most favored activity. However, Patterning Activity and Monkeys in the House (Operations) were mentioned six times in the top 3. Number Search was listed five times. Shape Search/Sort and Constructing Shapes were each mentioned three times. Measuring Worms, Freeze Game and Shape Pattern Blocks were mentioned twice. And Dot Fun and Number Music were each listed once. Comparing Blocks was the only game that did not appear in any one of the teachers' top three favorite activities.

Teacher confidence. Compared to pre-coaching, a majority of teachers (seven) reported that their confidence in teaching math "improved significantly." Two teachers reported that their confidence in teaching math "improved a little bit," and one teacher reported that her confidence had "not changed."

Child learning gains. As part of the weekly activity logs, teachers documented any learning gains they observed in specific students during the activity. Below are some examples of learning gains that teachers reported for children:

> *"Child X now is really recognizing numbers 2, 3, and 4!"*
>
> *"I was surprised I thought that Child X would not be able to remember how many monkeys, but today he finally got the hang of it and could keep track when we did the adding one"*
>
> *"I am so impressed with my class and freezing at the number. Almost all of them can do it now and now we can all move on to the counting-on level as a class"*
>
> *"Child X, Child Y and Child Z are especially learning their shapes now. By the end of small-group they all could recognize difference between triangle and rectangle"*
>
> *"The patterning templates are so great because I can make note of where a child left off and then I can start with that template next time. Three of my students made it to the highest level and most of my students can do the patterns without relying on the colored boxes now"*
>
> *"It's interesting how Child X can do the measurement activity so well now- and he is my youngest, but then he is not interested in learning his numbers. And then with Child Y, she has improved in her numbers and knowing how many, but this measuring activity is not clicking for her."*
>
> *"I am mainly just so excited that so many more of my students understand the one-to-one that Teacher Lillie showed me. We have been practicing putting our fingers on the materials when we count, and children are understanding more the "how many?" question"*

Post-Hoc Analyses

Given findings from Study 1, I was interested in whether there were associations between fidelity of implementation (observation), teachers' understanding of math development (teacher survey), and teachers' understanding of integration of EF in the activities (teacher survey). Even though my sample size was small, I conducted pairwise correlations to examine relations between teacher characteristics and elements of teacher understanding and fidelity of implementation (see Table IV.9). Correlations revealed that only teaching experience was statistically significantly correlated with teachers' understanding of math development post-coaching (r=.66), but teacher education (1= 2-year degree, 2= Bachelor's degree, 3= Master's degree) was not correlated with math development understanding (r=.08). Neither teaching experience nor education were significantly correlated with teachers' understanding of EF integration mid-coaching (r=-.04, r=.47), or post coaching (r=-.03, r=.36), nor fidelity of implementation with math (r=.33, r=.02) or EF integration (r=-.08, r=-.12). However, more educated teachers reported higher confidence in teaching math pre-coaching (r=.78) and reported that their confidence did not improve that much post-coaching (r=-.73).

Teachers who asked more questions about math throughout the intervention had longer coach meetings (r=.67), were rated as having a higher interest and more positive attitude in the beginning of the intervention (r=.79), and had a better understanding of math skill development post coaching (r=.70). Teachers who had higher fidelity of implementation on EF integration were also better at articulating their understanding of EF integration post-coaching (r=.75), but not mid-coaching (r=.11).

Discussion

Overall, teachers and coaches had a positive response to the content and coaching of the math/EF activities that spanned math topics from number sense, operations, patterning, shape recognition, and measurement integrated with directions that explicitly recruited children's executive functioning skills such as response inhibition, cognitive flexibility, and working memory. Most notably, after 16 weeks of in-classroom coaching, there was some evidence for differences between teachers' understanding of EF and math, and their implementation of the EF integration and math skill development portion of the activities. Furthermore, coaches appeared to understand the EF integration better than teachers, but coaches struggled more with articulating children's math skill development compared to teachers. This study may have implications for how the field educates teachers on math skill development, EF development, and the integration of the two in their classroom practice. This section discusses the specific strengths and weaknesses in the content and coaching of these activities and how to improve this intervention moving forward.

Feasibility of Math/EF Coaching Model

A primary aim of this study was to gather evidence for teachers' and coaches' response to the coaching model and intervention to inform alterations of this intervention moving forward. Overall, teacher interest and confidence in teaching math improved from pre to post intervention, which suggests that when this intervention is offered to more teachers, they might also respond as favorably. Another indication that this intervention is feasible in classrooms with teachers and coaches is that children were able to remain engaged in a classroom setting with teachers implementing the activities.

Interestingly, coaching seemed to benefit teachers with less education (they reported the largest gains in math teaching confidence pre to post). And more experienced teachers (more years of teaching preschool) made the most gains in math knowledge pre to post. This has important implications for coaching in the Los Angeles Head Start context where only 50% of teachers have a BA degree and average teaching experience is 11 years (Office of Head Start, 2019).

Despite the promising factors of this intervention, there are several improvements to consider future iterations of this intervention. One response from the teachers is that they wanted more introductory level activities and coaching for children of the lowest math skill. Teachers expressed that some of the children in their classroom were not progressing through levels and could barely pass level 1 of any of the activities. In *Learning and Teaching Early Math* (Clements & Sarama, 2014), the authors detailed how teaching math to the lowest skilled learners is the most difficult because it involves substantial repetition that does not immediately appear to be beneficial. Similar to how oral language input is important for infants to eventually produce their own language, the effects of math instruction may not be observed right away for children who enter the classroom with little math experience because a foundation has to be built before children's math skills can begin to improve. Teachers should be encouraged to continue to provide repetition of math teaching and math exposure to young children even when an immediate impact is not observed in a child's knowledge. In future iterations of this coaching framework, I will introduce more daily introductory practices for teachers to implement with children who cannot recognize numerals or demonstrate one-to-one correspondence.

A critique from the coaches was that they wanted the activities to be more explicitly aligned with existing tools that they used such as the Classroom Assessment Scoring System (CLASS) and Early Childhood Environment Rating Scale (ECERS). And teachers additionally expressed wanting guidance on how the activities could help them complete children's developmental profiles in the Desired Results Developmental Profile (DRDP). I am already in the process of aligning math/EF activities with these assessment and quality tools that teachers and coaches use as part of a partnership with Child 360. For example, when Child 360 coaches "coach on the CLASS" they typically isolate a specific component (e.g., Concept Development, a dimension of Instructional Support) and target different indicators within this component such as Analysis & Reasoning, Creating, Integration and Connections to Real World. Within these indicators, coaches provide teachers with specific strategies they can use such as prediction/experimentation, classification/comparison, integration with previous knowledge, child producing own work, why and/or how questions, and others. So, to integrate math/EF content with this "CLASS coaching," I connected my math/EF activities to these specific strategies. For example, to target geometry, I highlighted analysis & reasoning (indicator) and the classification/comparison (strategy) for my Shape Sort activity where children have to classify different shapes by a specific attribute, and then this attribute switches. Another example is in the creating (indicator) using the planning (strategy) for the Measuring Worms and Shape Building activities where children have to first pick out the materials they will be using (e.g., long vs. short sides and how many; and using short vs. long measuring units) before building and measuring. The hope is that this alignment with CLASS coaching will make teachers and coaches more engaged and likely to uptake the activities to improve upon their math instructional knowledge.

Teacher Understanding of Math and EF

Although not a primary aim, this study also offers preliminary evidence of the relation between teacher understanding of math and EF and their fidelity of implementation. All ten teachers demonstrated an increase in their understanding of math skill development from mid-coaching to post-coaching. Post-coaching, teachers scored the highest on their articulation of patterning development and geometry development, specifically, and struggled more with explaining the development of number sense and operations. Number sense includes more subskills than patterning and geometry (Clements & Sarama, 2004), which teachers could have mixed up in their articulation of math skill development (e.g., numeral recognition vs. recite counting vs. cardinality). Yet despite the overall increase in teacher math knowledge, only five teachers were able to implement the math skill level of a randomly selected math/EF activity to adequate fidelity. This aligns with past work on teachers' difficulty with identifying children's math skills and practicing appropriate developmental sensitivity to individual skills (Chen, McCory, Adams & Leow, 2014). In another math intervention with scripted activities following a specific sequence of math skills (High 5's), facilitators also struggled the most with differentiating instruction to individual children of different skill levels (Jacob, Erickson, & Mattera, 2018). It could be the case that over time teachers' practice will catch up to their increased knowledge of math, as has been demonstrated by increased fidelity over time with the Building Blocks curriculum (Clements & Sarama, 2007; Sarama, Clements, Starkey, Klein, & Wakely, 2008). Typically, interventions are evaluated after teachers have practiced the curriculum for one year (Clements & Sarama, 2007; Mattera, Jacob, & Morris, 2018; Presser, Clements, Ginsburg & Ertle, 2015). Teachers in this study learned 12 activities over the course of 16 weeks before a randomly selected activity was observed for fidelity, so it is possible that the observed activity was only practiced by the teacher once or twice on their own, so fidelity scores should be interpreted with caution.

Interestingly, teachers' understanding of math development and fidelity to the math skill levels in the activities were related to teachers having a more positive attitude and interest from the beginning of the intervention, teachers having more years of experience teaching preschool, and teachers asking more questions about children's math development at the coach meetings. This aligns with evidence from the Building Blocks curriculum, where more experienced teachers also achieved higher levels of fidelity more quickly than less experienced teachers

(Clements, Sarama, Wolfe, & Spitler, 2014). When it comes to teacher understanding, however, despite teacher understanding of the curriculum being a part of the theory of change (for example, see TRIAD model; Sarama, Clements, Wolfe, & Spitler (2016)), there has been historically no measure for “teacher understanding” beyond teachers’ fidelity of implementation scores in the study of preschool math curricula. As such, it is difficult to compare how teachers’ understanding of math in this math/EF intervention compares to that of other work.

Teachers made less improvement in articulating how EF was integrated into each of the math activities, and their understanding of EF was relatively low at both mid-coaching and post-coaching time points. However, despite their inability to articulate EF, most teachers demonstrated adequate fidelity to the EF component of the math/EF activities (N=8 teachers achieved this). This relates to past work documenting how elementary school teachers in general lack an understanding of EF development (Little, 2017), despite acknowledging the importance of EF skills for learning (Nyroos, Wiklund-Hornqvist, & Lofgren, 2018), but have difficulty explaining EF. Similarly, this inability of teachers to future grasp EF development contributes to the puzzle of EF interventions’ historic lack of impact on children’s EF skills, despite teachers’ high levels of fidelity of implementation. It could be simple for teachers to mimic the different strategies that promote children’s EF, yet not have a clear understanding of this phenomenon, which could make it difficult for this practice to be sustained. For instance, Wasik & Hindman (2012) have illustrated the importance of teachers’ literacy knowledge for the sustained fidelity of a literacy intervention. Yet the field has not had as much opportunity to assess this in the case of EF interventions because studies that have demonstrated impacts on children’s EF skills have been implemented by researchers and not teachers (Tominey & McClelland, 2011; Rabiner et al., 2010). In the current study, EF was inherently exercised by the nature of the activity setup and “rules,” so it is possible that teachers did not have to be as overtly focused on this, nor deeply understand it in the same way they had to for math. Future work should continue to measure “EF understanding” and how it contributes to EF intervention success.

Limitations

Overall, there were several limitations of this study. For one, despite the opportunity to collect a rich amount of data, activities were implemented with only 10 teachers and four coaches, which limits the external validity of the findings and the statistical power of the study. A second serious limitation was the lack of coach time devoted classroom visits. Child 360 funded their attendance to the training, but beyond that, I provided compensation for the survey they completed, every time they coded a teacher survey/fidelity observation, and the interview they participated in. But motivating coaches to complete each one of these tasks proved to be very difficult. Coaches were only able to visit and model for teachers twice a month instead of once a week, and some coaches did not even visit in December or January at all due to their other coaching paperwork and deadlines. I think that in a future iteration of this work, I will set aside more funding for coaches to compensate them for their time in the classrooms because I do not think they had enough incentive to visit classrooms more frequently. Additionally, I think a broader limitation of working with Child 360 coaches was that their teacher coach ratio was rather high (26:1) and as a result, they could not possible give as much time to each teacher that would be needed, which points to a fundamental issue working in this context. A third limitation of this work was that I created my own measures. Although influenced by other researchers’ measures, this still limited my ability to connect this research to other work. A fourth limitation was only collecting one fidelity of implementation observation per teacher, when I was planning to collect five (due to COVID-19). However, I was fortunate enough that I had a 100%

participation rate in the virtual survey and interviews from the teachers and coaches. Lastly, this study did not provide any evidence for whether these activities actually impacted children's EF and math skills.

Future Directions

Currently, I am in the midst of Year two of this study to implement improvements to the coaching framework and have created an official math training series through Child 360. I have taken results from this study to modify the coach training to connect it explicitly to coaches' existing practices and tools they use (e.g., CLASS, ECERS- examples I gave earlier) and provide a toolkit for teachers that aligns math/EF activities with the DRDP. A total of 33 coaches participated in the coach training in early October 2020 and have started using this integration of the math/EF content with the CLASS and DRDP as part of their coaching. As soon as COVID-19 restrictions are lifted, I am going to continue to build out and pilot this intervention with a larger sample and evaluate its impact on children's math and EF skills.

Conclusion

This study highlighted the feasibility of a combined math/EF intervention in the context of preschool classrooms with teachers and coaches. Findings revealed four major contribution to the literature. One, a sample of teachers and coaches achieved fidelity of implementation on one randomly selected math/EF activity, which is encouraging for the promise of these activities with more teachers and coaches. Two, teachers increased in their math teaching confidence and understanding of math development across the intervention period, and there were some notable relations between these factors and teachers' level of education and fidelity of implementation scores. Three, teachers struggled with articulating how executive function skills were exercised within the math/EF activities, which has implications for how we coach teachers on children's

EF development in the future. And lastly, this study shed light on some major barriers with continued math/EF coaching (and domain-specific coaching more generally) in Los Angeles Head Start centers due to the competing coaching efforts and goals of the Classroom Assessment Scoring System.

References

Arnold, D.H., Fisher, P.H., Doctoroff, G.L., & Dobbs, J. (2002). Accelerating math development in Head Start classrooms. *Journal of Educational Psychology, 94*(4), 762-770.

Baddeley, A. (1992). Working Memory. *Science, 255*(5044).

Barkley, R.A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94.

Barnett, W.S., Jung, K., Yarosz, D.J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effets of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*(3), 299-313.

Barrett, D. (2017). *STEAM framework feasibility study*. Los Angeles: Los Angeles Universal Preschool. Available at: https://child360.org/wpcontent/uploads/2017/06/LAUP_CA_FULLREPORT_STEAMFFS_rev20170619.pdf

Bay-Williams, J.M., & McGatha, M.B. (2014). *Mathematics coaching: Resources and tools for coaches and leaders, K—12*. Boston, MA: Pearson.

Best, J.R., Miller, P.H., & Naglieri, J.A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327-336.

Biancarosa, G., & Bryk, A.S. (2011). Efficacy of literacy collaborative professional development: A summary of findings. *Journal of Reading Recovery, 10*(2), 25-32.

Bierman, K.L., Domitrovich, C.E., Nix, R.L., Gest, S.D., Welsch, J.A., Greenberg, M.T., Blair, C., Nelson, K.E., & Gill, S. (2008). Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program. *Child Development, 79*(6), 1802-1817.

Blair. C., & Razza, R.P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647-663.

Blair. C., Protzko, J., & Ursache, A. (2011). Self-regulation and the development of early literacy. In D. Dickinson & S Neuman (Eds.), *Handbook of early literacy research* (pp. 20-35). New York: Guilford Press.

Blair, C., Ursache, A., Greenberg, M., Vernon-Feagans, L., & The Family Life Project Investigators. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter-word knowledge in the early elementary grades. *Developmental Psychology, 51*(4), 459-472.

Bloom, H., & Weiland, C. (2015). Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the national Head Start Impact Study (Working Paper). Retrieved from http://www.mdrc.org/sites/default/files/quantifying_variation_in_head_ start.pdf

Bock, A., Cartwrigth, K.B., Gonzalez, C., O'Brien, S., Robinson, M.F., Schmerold, K., Shriver, A., & Pasnak, R. (2015). The role of cognitive flexibility in pattern understanding. *Journal of Education and Human Development, 4*(1), 19-25.

Bodrova, E., & Leong, D.J. (2007). Play and Early Literacy: A Vygotskian Approach. From Play in Literacy in Early Childhood: Research From Multiple Perspectives.

Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology, 19*(3), 273-293.

Bull, R., Epsy, K.A., & Wiebe, S.A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205-228.

Campbell, P.F., & Malkus, N.N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111*(3), 430-454.

Cameron, C.E., Connor, C.M., Morrison, F.J., & Jewkes, A.M. (2008). Effects of classroom organization on letter-word reading in first grade. *Journal of School Psychology, 46*(2), 173-192.

Cameron, C.E., & Morrison, F.J. (2011). Teacher Activity Orienting Predicts Preschoolers' Academic and Self-Regulatory Skills. *Early Education and Development, 22*(4), 620-648.

Capraro, R.M., Capraro, M.M., & Rupley, W.H. (2011). Reading-enhanced problem solving: A theoretical model. *European Journal of Psychology of Education, 27*(1), 91-114.

Carter, D.R., & Norman, R.K. (2010). Class-wide positive behavior support in preschool: Improving teacher implementation through consultation. *Early Childhood Education Journal, 38*(4), 279-288.

Chen, J., McCray, J., Adams, M., & Leow, C. (2014). A survey study of early childhood teachers' beliefs and confidence about teaching early math. *Early Childhood Education Journal, 42*(6), 367-377.

Choi, J.Y., Castle, S., Williamson, A.C., Young, E., Worley, L…Horm, D.M. (2016). Teacher-child interactions and the development of executive function in preschool-aged children attending head start. *Early Education and Development*, *6*, 751-769.

Copley, J. V. (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and kindergarten mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Clark, C.A.C., Pritchard, V.E., & Woodward, L.J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176-1191.

Clements, D.H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks Project. *Journal for Research in Mathematics Education, 38*(2), 136-163.

Clements, D. H., Sarama, J., Unlu, F., & Layzer, C. (2012, March). *The efficacy of an intervention synthesizing scaffolding designed to promote self-regulation with an early mathematics curriculum: Effects on executive function.* Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Clements, D.H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36*, 79-90.

Cohen, E., & Kaufmann, R. (2000). Early childhood mental health consultation. Washington, DC: Center for Mental Health Services, SAMHSA, U.S. Department of Health and Human Services.

Darrow, C.L. (2013). The effectiveness and precision of intervention fidelity measures in preschool intervention research, 24(8), 1137-1160.

DeFlorio, L., & Beliakoff, A. (2015). Socioeconomic status and preschoolers' mathematical knowledge: The contribution of home activities and parent beliefs. *Early Education and Development*, *26*(3), 319–341.

Diamond, A. (2016). Why improving and assessing executive function early in life is critical. From Executive Function in Preschool-Aged Children: Integrating Measurement, Neurodevelopment, and Translational Research. J.A Griffith, P. McCardle, and L.S Freund (Editors).

Diaz, G., & McClelland, M. M. (2018). Latino families and schools: Identifying challenges and highlighting strengths to support children's early learning. In S. Sonnenschein & B. Sawyer (Eds.), *Academic socialization of young Black and Latino children: Building on family strengths*. New York: Springer.

Duda, M.A., Dunlap, G., Fox, L. (2004). An experimental evaluation of positive behavior support in a community preschool program. *Topics in Early Childhood Special Education, 24*(3), 143-155.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L., Feinstein…(2007). School readiness and later achievement. *Development Psychology, 43*(6), 1428-1446.

Erchick, D., Brosnan, P., Forrest, D., Douglass, L., Grant, M., & Hughes, K. (2007, March). *Findings from the first year of a K—6 mathematics coaching project*. Paper presented at the Research Presession of the National Council of Teachers of Mathematics, Atlanta, GA.

Evans, G.W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly, 23*(4), 504-514.

Fantuzzo, J.W., Gadsden, V.L., & McDermott, P.A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal, 48*(3), 763-793.

Farran, D. C., Lipsey, M. W., & Wilson, S. J. (2011, November). *In Curriculum and pedagogy: Effective math instruction and curricula*. Paper presented at the early childhood math conference, Berkeley, CA.

Farran, D., & Wilson, S. (2014). *Achievement and self-regulation in pre-kindergarten classrooms: Effects of the tools of the mind curriculum*. Peabody Research Institute. Retrieved February 2016, from http://www.executivefunctionthinktank.com/uploads/2/0/9/8/2098082/achievement_and_self_regulation.pdf

Farran, D. C., Silveri, B., & Culp, A. (1991). Public preschools and the disadvantaged. In L. Rescorla, M. C. Hyson, & K. Hirsh-Pasek (Eds.), *Academic instruction in early childhood: Challenge or pressure?* (pp. 65–73). San Francisco: Jossey-Bass.

Farran, D.C., Wilson, S.J., & Lipsey, M. (2013). Effects of a curricular attempt to improve self-regulation and achievement in prekindergarten children. *Society for Research on Educational Effectiveness.*

Filcheck, H.A., McNeil, C.B., Greco, L.A., & Bernard, R.S. (2004). Using a whole-class token economy and coaching of teacher skills in a preschool classroom to manage disruptive behavior. *Psychology in the Schools, 41*(3), 351-361.

Fuhs, M., & McNeil, N.M. (2013). ANS acuity and mathematics ability in preschoolers from low income homes: contributions of inhibitory control. *Developmental Science, 16*(1), 136-148.

Gersten, R., Chard, D., & Baker, S. (2000). Factors Enhancing Sustained Use of Research-Based Instructional Practices. *Journal of Learning Disabilities, 22*(5), 445-4567.

Ginsburg, H. P., Klein, A., & Starkey, P., (1998). The development of children's mathematical thinking: connecting research with practice. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of child psychology: Child psychology in practice* (5th ed., Vol. 4, pp. 401–476). New York: Wiley.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496-499.

Griffin, S. (2003). Number Worlds: A research-based mathematical program for young children. In D.H. Clements and A. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 325-342). Mahweh, NJ: Lawrence Erlbaum Associates.

Hindman, A. H. (2013). Mathematics instruction in Head Start: Nature, extent, and contributions to children's learning. *Journal of Applied Developmental Psychology, 34*(5), 230-240.

Hsieh, W., Hemmeter, M.L., McCollum, J.A., & Ostrosky, M.M. (2009). Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly, 24*(3), 229-247.

Hughes, C., & Cline, T. (2014). An evaluation of the preschool PATHS curriculum on the development of preschool children. *Educational Psychology in Practice, 31*, 73–85.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036.

Hulleman, C.S., & Cordray, D.S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-110.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257-1274.

Huzinga, M., Dolan, C.V., & van der Molen, M.W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036

Iseman, J.S., & Naglieri, J.A. (2011). A cognitive strategy instruction to improve math calculation for children with ADHD and LD: a randomized controlled study. *Journal of Learning Disabilities, 44*(2), 184-195.

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*(4), 512-552.

Jacoby, J. W., & Lesaux, N. K. (2017). Policy-based instructional reform in early education: How U.S. Head Start teachers perceive instructional mandates. *International Journal of Child Care and Education Policy, 11*(9).

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E., Clements, D., Sarama, J., & Duncan, G. J. (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *Journal of Research on Educational Effectiveness, 11*, 339–374.

Joswick, C., Clements, D.H., Sarama, J., Banse, H., & Day-Hess, C.A. (2019). Double Impact: Mathematics and Executive Function. *Teaching Children Mathematics, 25*(7), 416-426.

Keena, L., Conroy, S., & O'Sullivan, A., & Downes, M. (2019). Executive functioning in the classroom: Primary school teachers' experiences of neuropsychological issues and reports. *Teaching and Teacher Education, 86*.

Keenan, L., O'Sullivan, A., & Downes, M. (2020). Teachers' experiences and understanding of executive functions in Irish primary school classrooms: Findings from a mixed-methods questionnaire. *Irish Educational Studies*

Kyger, M., & Barnhart, M. (2017). *Effects of high-quality preschool on early literacy and mathematics skills—A regression-discontinuity approach*. Los Angeles: Los Angeles Universal Preschool. Available at: http://child360.org/wpcontent/uploads/2017/02/LAUP_CDSR_FULLREPORT_REGDISC_rev20170224.pdf

Lefevre, J. A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, *4,* 1-9.

Lipsey, M.W., Nesbitt, K.T., Farran, D.C., Dong, N., Fuhs, M.W., & Wilson, S.J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevant of selected measures. *Journal of Educational Psychology, 109*(8), 1084-1102

Little, M. (2017). Racial and Socioeconomic Gaps in Executive Function Skills in Early Elementary School: Nationally Representative Evidence From the ECLS-K: 2011. *Educational Researcher, 46*(2), 103-109.

Lloyd, C. M., & Modlin, E. L. (2012). *Coaching as a key component in teachers' professional development: Improving classroom practices in head start settings* (OPRE Report 2012-4). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families.

Lopez, E.M., Gallimore, R., Garnier, H., & Resse, L. (2007). Preschool antecedents of mathematics achievement of Latinos: The influence of family resources, early literacy experiences, and preschool attendance. *Hispanic Journal of Behavioral Sciences, 29*(4), 456-471.

Los Angeles County Office of Education (LACOE) Head Start-State Preschool Division. (2016). *Annual Report*.

Mattera, S., Lloyd, C. M., Fishman, M., & Bangser, M. (2013). *A first look at the Head Start CARES Demonstration: Large-scale implementation of programs to improve children's social-emotional competence* (OPRE Report 2013-47). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

McClelland, M.M., Cameron, C.E., McDonald, C., Farris, C.L., Jewkes… Morrison, F.J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947-959.

McClelland, M.M., & Cameron, C.E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development, 133,* 29-44.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Frontiers in Psychology, 5*, 1-14.

Meador, D.N., Turner, K.A., Lipsey, M.W., & Farran, D.C. (2013). *Administering measures from the PRI learning-related cognitive self-regulation study*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology, 7*.

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making pre-K count: Improving math instruction in New York City*. New York, NY: MDRC.

Neuman, S., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532-566.

Nesbitt, K.T., Farran, D.C., & Fuhs, M.W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology, 51*(7), 865-878.

Nguyen, T., & Duncan, G.J. (2019). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly, 46*(1), 49-61.

Nyroos, M., Wiklund-Hornqvist, C., & Lofgren, K. (2018). Executive function skills and their importance in education: Swedish teachers’ perceptions. *Thinking Skills and Creativity, 27*, 1-12.

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.

Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., LoCasale-Courch, J…Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*, 956-975.

Piasta, S.B., Logan, J.A.R., Pelatti, C.Y., & Capps, J.L. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms

Piasta, S.B., Justice, L.M., O'Connell, A.A., Mauck, S.A., Weber-Mayrer, M., Schachter, R.E., Farley, K.S., & Spear, C.F. (2017) Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Educational Effectiveness, 10*(2), 354-378

Ponitz, C.C., Rimm-Kaufman, S.E., Brock, L.L., & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade.

Prager, E.O., Sera, M.D., & Carlson, S.M. (2016). Executive function and magnitude skills in preschool children. *Journal of Experimental Child Psychology, 147,* 126-139.

Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care: Analyzing qualitative data. *British Medical Journal, 320*, 114–116.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology, 102*(2), 299-312.

Preschool Curriculum Evaluation Research Consortium (2008). *Effects of preschool curriculum programs on school readiness* (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S Department of Education.

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. *The Elementary School Journal*, *96*(4), 363-384.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K., & Sardin, L. (2008). Improving pre-school classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10–26.

Raver, C.C., McCoy, D. C., Lowenstein, A. E., & Pess, R. (2013). Predicting individual differences in low-income children's executive control from early to middle childhood. *Developmental Science*, *16*(3), 394–408.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity? Rising inequality and the uncertain life chances of low-income children* (pp. 91–116). New York, NY: Russell Sage Foundation Press.

Reinke, W.M., Herman, K.C., & Dong, N. (2018). The Incredible Years Teacher Classroom Management Program: Outcomes from a Group Randomized Trial. *Prevention Science, 19*(8), 1043-1054.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*(4), 958-972.

Rudd, L. C., Lambert, M. C., Satterwhite, M., & Smith, C. H. (2009). Professional development + coaching = enhanced teaching: Increasing usage of math mediated language in preschool classrooms. *Early Childhood Education Journal*, *37*(1), 63–69.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Schmitt, S.A., McClelland, M.M., Tominey, S.L., & Acock, A.C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly, 30*, 20-31.

Sheridan, S.M., Edwards C.P., Marvin, C.A., & Knoche, L.L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education and Development, 20*(3), 377-401.

Siegler, R.S. (2009). Implications of cognitive science research for mathematics education. In Kilpatrick, J., Martin, W.B., & Schifter, D.E. (Eds.), *A research companion to principles and standards for school mathematics* (pp. 219-233). Reston, VA: National Council of Teachers of Mathematics.

Simmons, F.R., Willis, C., & Adams, A. (2012). Difference components of working memory have different relationships with difference mathematical skills. *Journal of Experimental Child Psychology, 111*(2), 139-155.

Sophian, C. (2004). Mathematics for the future: Developing a Head Start curriculum to support mathematics learning. *Early Childhood Research Quarterly, 19*(1), 59-81.

Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on* *mathematics in early childhood education* (pp. 253–276). Charlotte, NC: Information Age.

Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America, 110*(45), 18116-18120.

Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: Differential developmental gateways to literacy and numeracy. *Child Development, 83*(6), 2028-2041.

Stein, M.L., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Yes, L…Compton, D.L. (2008). Scaling up and early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis, 30*(4), 368-388.

Tominey, S.L., & McClelland, M.M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education and Development, 22(*3), 489-519.

Tout, K., Isner, T., & Zaslow, M. (2011). *Coaching for quality improvement: Lessons learned from quality rating and improvement systems.* Retrieved from http://www.childtrends.org/wpcontent/uploads/2013/05/201135CoachingQualityImprovement1.pdf

Usai, M.C., Viterbori, P., & Traverso, L. (2018). Preschool executive function profiles: Implications for math achievement in grades 1 and 3. *Journal of Research in Childhood Education, 32*(4), 404-418.

Van der Ven, S.H., Kroesbergen, E.H., Boom, J., & Leseman, P.P.M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*(1), 100-119.

Vygotsky, L.S. (1962). Thought and language. Cambridge, MA. MIT Press.

Wanless, S. B., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2011). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Education & Development, 22,* 461–488

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455-469.

Wasik, B.A., Bond M.A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*(1), 63-74.

Weiland, C., Wolfe, C. B., Hurwitz, M. D., Clements, D. H., Sarama, J. H., & Yoshikawa, H. (2012). Early mathematics assessment: Validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology, 32*(3), 311–333.

Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring development of executive function skills and receptive vocabulary in preschool-aged children: A look at the direction of the developmental pathways. *Infant and Child Development, 23*(1), 4–21.

Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open, 4*(1), 1-16.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting* *and Clinical Psychology, 65*(1), 93-109.

Webster-Stratton, C., & Reid, M. J. (2003). The incredible years parents, teachers and children training series: A multifaceted treatment approach for young children with conduct problems. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 224-240). New York, NY, US: The Guilford Press

Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In D. L. Molfese & V. J. Molfese (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139-185). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Whitebrooke, M., Bellm, C., & Schaak, D. (2013). *Supporting teachers as learners: A guided for mentors and coaches in early care and education*. Washington, D.C.: American Federation of Teachers.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* Washington, DC and New York, NY: Society for Research in Child Development and Foundation for Child Development.

Yoshikawa, H., Leyva, D., Snow, C.E., Trevino, E., Barata, M., Weiland, C., Gomez, C.J., Moreno, L., Rolla, A., D'Sa, N., Arbour, M.C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology,* 51(3), 309.

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.

Figure 1V. Study III Design Overview

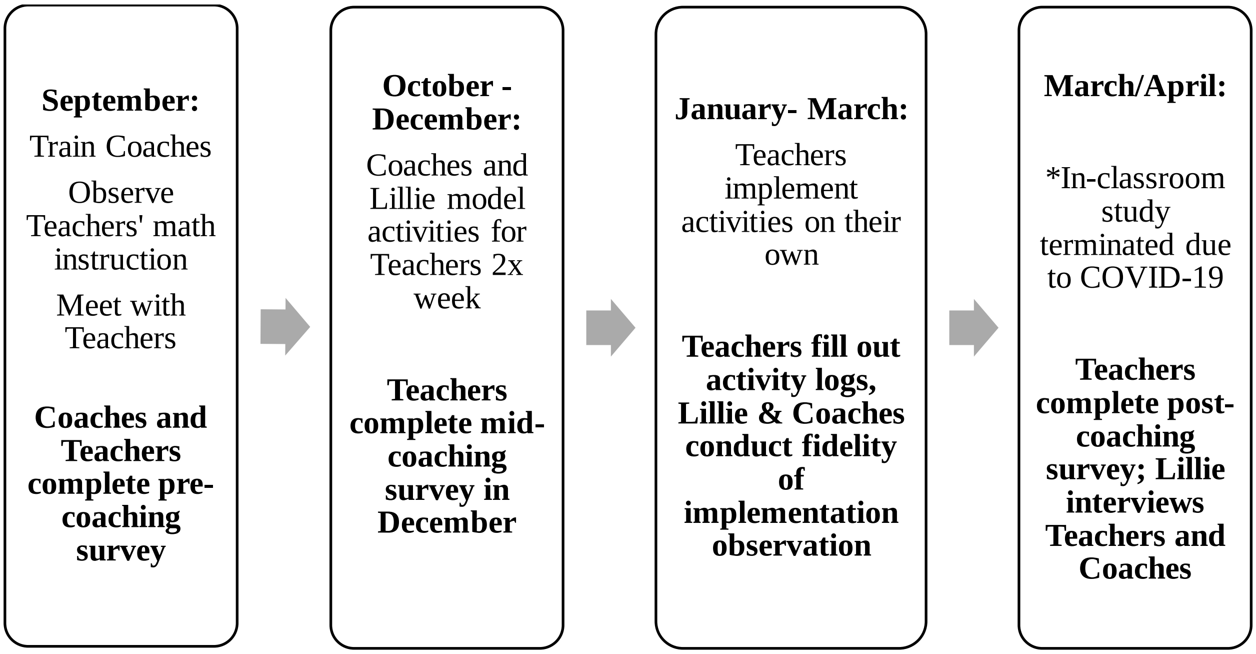
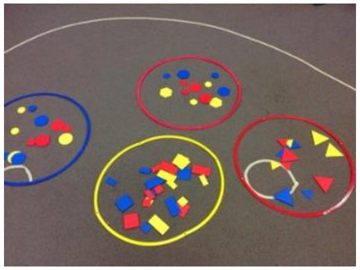


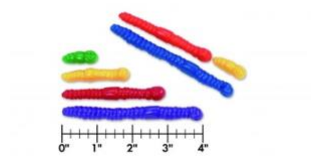
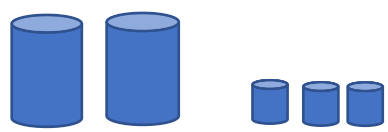
Table IV.1. Overview of Math/EF Activities

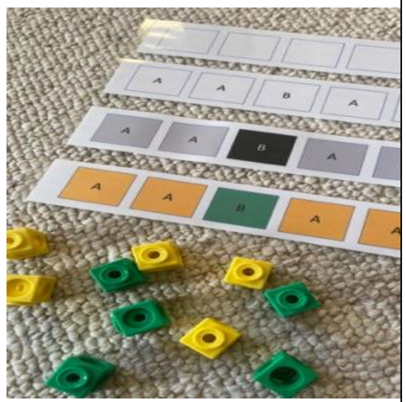
Activity Name	Math topic	Summary/Materials	Executive functioning	Developmental sequence
Number Search (Large or small group)	Number recognition	Searching for a specific number in a display of numbers; Materials: Numbered cards 1-10 in blue, green, red and small and big sizes. 3 3 2 4 4 3	Cognitive flexibility; Looking for all the "3's", and then the "green 3's"	Numerals 1-5 Find a certain size OR color number "big 2's", or "blue 2's" Numerals 1-10 Find a specific size AND color "small blue 4's"
Freeze Game (Large group or transition)	Counting; Counting on; Cardinality	Crouch body down and inch upwards until standing while counting 1-10 Materials: None! 	Inhibitory control; Freeze body at a specific number	Freeze at numbers 4-6 Freeze at numbers 7-10 or 1-3 Freeze at a number and then "count on" to another number; e.g., Freeze at 3 and then count on to 5.
Monkeys in the House (Small group)	One-to-one correspondence; Cardinality; Addition and subtraction	Two houses and monkey counters; Monkeys are placed in and out of the houses; teacher asks how many monkeys are in the house now	Working memory; Have to keep track of how many monkeys are in the house (no more than 4 monkeys in one house)	Use only one house; add one monkey at a time. Use two houses; add one monkey at a time.

		Materials: Houses, monkey counters (or other animal), and magnetic ten-frame. 		Use two houses; add two monkeys or take one monkey away Use two houses; take one monkey from one house and add to other. **** For all levels, there are never more than four monkeys in a house
Dot fun (Large or small group; Transition)	One-to-one correspondence; Subitizing; Magnitude comparison	Counting dots on cards depending on what children are asked to search for; some of the cards are flashed quickly so children must focus, and find the "speed" component engaging Materials: Dot cards; Some with all the same color/size dots; Some with different color and sized dots	Cognitive flexibility; Have to count only the specific color dot asked for	***Asking "How many?", is a question a child should be asked every day—to work on knowing that NUMBER is not just counting, but actually has meaning (e.g., cardinality)*** How many dots do you see? No more than 3- all dots are same color Different colored dots; How many (specific color) dots do you see Different colored and sized dots; How many (specific color) dots do you see; How many (specific size) dots? Now, how many in all?

				Which color dots has MORE? How do you know?
Shape Search (Large or small group) Shape Sort (Large or small group)	Shape recognition; Shape attributes	Children search for a shape: “green triangles”, or just “triangles”, or “big triangles” etc. Children then sort the shapes. By color, size, type, and number of sides Materials: Green, blue, red, shape cards including circles, triangles, rectangles, squares, trapezoids, rhombuses, hexagons 	Cognitive flexibility; Similar to number search: switch the rule of what they are searching or sorting (sort by color, sort by shape; find all	Searching for triangles Searching for rectangles Then add a specific color or size triangle/rectangle Add squares, trapezoids, and rhombuses Add hexagon ----------------------- For sorting part of this activity: First sort by color (red shapes in one pile, green in another etc.) Then sort by type of shape (e.g., triangles in one pile, rectangles in another, trapezoids in another) Then sort by number of sides (3 sides vs. 4 sides; triangles in one pile, rectangles AND trapezoids in another)
Constructing shapes (Small group)	Shape recognition; “Shorter” and “longer”; Shape attributes	Building triangles, squares and rectangles; emphasis on short and long sides, and corners	Cognitive flexibility; e.g., Have to make a triangle four different ways etc.	Making triangle with all the same length sides Making triangle in a different way (2 long and 1 short side)

		Materials: Magnetic sides (short and long) and corner balls to connect them 		Making a square (with all same sides) Making a rectangle with two long and two short sides
Number Music (Large group)	Number recognition	Walk around in a circle with numbers on the carpet; Music plays and when it stops, find a number to stand on, pick it up and try to match number with a peer's number. Materials: Number stickies (that stick to carpet) 	Inhibitory control; Freeze when the music stops & keep directions in mind	Numerals 1-5 Numerals 6-10
Measuring worms (Large and small group)	Measurement; Magnitude comparison; "Longest" and "shortest"; Ordering by length	Ready aloud book about an inchworm who measures animals ("Inch by Inch"); Then use measuring worm counters to measure animals at small group	Cognitive flexibility; Use little inch worms to measure; then use longer inch worms to measure the same animals	Measure up to 4 -inch animal only using small (1 inch) measuring worms Measure up to 6-inch animals using small (1 inch) measuring worms and then measure same

		Materials: "Inch by Inch" storybook; Measuring worm counters (1-inch, 2-inch, and 3-inch) 		animal with (2-inch) measuring worms; Why does it take fewer longer worms to measure the same thing? Measure two animals and say which one is longer, which one is shorter. How do you know?
Who has more? Comparing Blocks (Large or small group; Transition)	Magnitude comparison; One-to-one; Cardinality; Addition and subtraction	Give different sized items to two children and ask them to tell you which of them has more Materials: Use any toys/objects in the classroom 	Cognitive flexibility; Children focus on number of items, not the size of the items	Use up to 4 objects total Use up to 6 objects total How many more toys should John get so that him and Elle have the SAME?
Pattern Templates (Small group)	Patterning	Children use different patterning templates depending on their level and copy, extend or create a pattern with colored counters. Learn term "AB" etc. patterns	Cognitive flexibility; Switch between AB patterning, AAB patterning, and patterning using different colors etc.	Copy an AB pattern with exact colors Extend an AB pattern Copy an AAB or ABB pattern with exact colors

		Materials: Pattern templates; different colored counters or linking cubes (e.g., fruit counters, vehicle counters, monkey counters etc.) 		Create an AB, AAB, and ABB pattern with no color template (just black and white colors, or just the “AB” letters) Extend an AAB or ABB pattern Repeat above but with ABC patterns Create an AB, AAB, ABB, ABC pattern with different colors Create an AB, AAB, ABB, ABC pattern with no template (just are told to create “AAB pattern” Create own pattern *Can also have children move away from patterning by color and have them pattern by type of counter, ignoring the color “fruit counter, bear counter, fruit counter, bear… etc.”
Shape Picture Blocks (Small group)	Shape recognition; Spatial awareness; Fine motor skills	Children use fine motor skills to position shapes blocks on the cards.	Inhibitory control; Have to ignore color of block and focus on shape	Use the same color shape blocks as the shape pictures

		Materials: Pattern/shape cards (color and no color), original pattern/shape blocks (color e.g., yellow hexagon, green triangle, orange square), and different pattern/shape blocks (color e.g., red hexagon, purple square etc.) 		Use a different color shape block from the shape pictures Use a no color shape picture and mix of colored blocks

Table IV.2. Materials Teachers Had and/or Needed for the Intervention

All teachers had	Four teachers had, but rest of teachers did not have enough for small-group activity	Made for teachers (created, color printed, and laminated)	Purchased for teachers
Wooden Lakeshore houses (at least 2) Bear or Monkey Counters (at least 15)	More counters (at least 120); multi-colored (orange, purple, green, red, blue, yellow) Pattern Shape Block Cards (at least 15) Shape Pattern Blocks (at least 200)	Laminated number cards (big, small, red, green, and blue)- 60 cards per teacher Laminated dot cards- 25 cards per teacher Laminated pattern templates for different skill levels- 50 cards per teacher Laminated shape cards (big, small, hexagon, circle, trapezoid, rectangle, square, triangle, red, green, blue)- 100 cards per teacher Laminated animal cards of different lengths (2'' – 7''; tiger, elephant, frog, rabbit, mouse, bird)	Magnetic ten frames (2) with different colored circle magnets (10) "Inch by Inch" Storybook (1) Magnetic blocks for shape construction (1 100-pc set) Velcro numbers for carpet (3 sets of numbers 1-10) Shape Pattern Blocks (non-traditional colors e.g., red hexagon instead of yellow, purple triangle instead of green) (1 set of 100-pc)

Table IV.3. Observed Classroom Math Practices (September 2019; Pre-Coaching)

Math instructional instance	Math skill	N Teachers
Math at meal-time or brushing teeth	Counting out cups; Counting to 10	5
Math at small-group	Matching number of dots to numeral; using shape pattern blocks to create pictures; number tracing	4
Math at large-group	Counting attendance Identifying numbers on the calendar	5
Math at centers	Building with Magna tiles (2); Shape puzzles; Matching numbers with dots; Patterning with beads	5
Math during transition	"If you can find the 3, then you can line up" Washing hands for 10 seconds	3
Missed math instructional opportunity/developmentally inappropriate	"9 plus 6 girls is how many?"	1
	Reciting 1-30 and asking children to recognize the numbers 1-30	2
	Copying patterns but not identifying what part is repeating	2
	Working with quantities above 5 when some children cannot count to 5	7
	Using shape blocks without taking opportunity to teach the names of shapes, or practice naming	2
	"Why do you think?" and "How do you know?" questions that were too difficult (ex. "Why do you think this shape fits in this spot, and not this shape?"; "How do you know there are more girls than boys today?")	4

Table IV.4. Teachers' Fidelity of Implementation Scores

	Activity	EF	Math Level	Followed Script	Organization	Child Engagement
Teacher 1	Number Search	3.0	2.3	3.0	3.0	3.0
Teacher 2	Shape Search	2.3	1.0	1.3	2.0	3.0
Teacher 3	Monkeys	2.6	2.3	3.0	3.0	3.0
Teacher 4	Monkeys	3.0	3.0	3.0	3.0	3.0
Teacher 5	Number Search	3.0	2.6	3.0	2.6	3.0
Teacher 6	Number Search	1.0	1.0	2.6	1.0	1.0
Teacher 7	Monkeys	3.0	3.0	2.6	2.6	2.3
Teacher 8	Shape Search	2.6	2.3	2.0	1.3	1.3
Teacher 8	Monkeys	2.3	2.0	1.0	2.6	2.0
Teacher 9	N/A					
Teacher 10	Number Search	2.6	3.0	2.3	3.0	2.0
Teacher 10	Shape Search	3.0	2.6	2.3	3.0	3.0

Note. Fidelity scores range 1-3. Nine teachers received at least one observation; two teachers had two observations. Averaged scores across two coaches and author (with the exception of math level, where reliability was not reached among 3 coders)

Table IV.5. Comparing Math/EF Activities (Post-Coaching Scores)

	Understanding of EF integration	Understanding of Math level	Teacher rated coaching received	% of Teachers reported did not need more coaching	No. times teachers mentioned in their top 3 favorite activities
Number Search	3.0	2.2	4.0	80%	4
Freeze Game	3.5	2.1	4.0	80%	2
Monkeys House	3.0	2.7	3.9	70%	5
Compare Blocks	1.8	2.1	2.9	60%	0
Dot Fun	2.6	2.9	3.6	80%	1
Shape Search	2.0	3.3	3.9	90%	1
Shape Sort	2.5	3.2	4.0	90%	2
Construct Shape	2.1	2.1	3.9	70%	3
Number Music	2.0	2.0	3.0	100%	1
Measure Worms	2.2	2.2	3.3	70%	2
Patterning	2.0	3.9	4.0	100%	5
Shape Picture	2.1	2.8	3.0	80%	2

Table IV.6. Emerging Themes from Teacher and Coach interview: Strengths & Weaknesses of Content & Coaching

	Strengths-Teachers (> 50% teachers mentioned)	Weaknesses-Teachers (> 50% teachers mentioned)	Strengths-Coaches (100% mentioned)	Weaknesses-Coaches (100% mentioned)
Content	Variety of materials helpful; Children very engaged in most activities; Breaking down math activity into specific skills is helpful	More materials for small-group; Not enough “introductory-level” activities—especially for geometry and measurement	Concrete, scripted activity booklet helpful for teachers to learn the content (and for us to coach them with)	Activities would benefit from being explicitly tied to goals of the CLASS, ECERS, and the DRDP
Coaching	Teachers liked the in-classroom modeling approach; Weekly coaching was favored by most (Four teachers preferred monthly or biweekly)	More coaching for how to support children of the lowest skill level; Coaching should integrate more with schedule of the day, curriculum and/or teachers’ existing lesson plan	Coach training itself was very informative; Modeling of activities is very helpful for teachers	Frequency of coaching needed to assist teachers with implementation is not realistic

Table IV.7. Teacher Reported Strengths & Weaknesses of Specific Math/EF Activities

	Strengths	Weaknesses
Number Search	"Good activity for number recognition and the switching between color and size is an interesting teaching strategy" "Combines number recognition with size and color which keeps children engaged"	"Stronger lamination of the number cards" "How do I help the children who really cannot recognize numbers?"
Freeze Game	"Children love this game" "Children always ask for this game"	"Some of the younger children not really internalizing the "counting" and can't stop at a number. How do I help them?" "Some children just fool around so the management can be difficult"
Monkeys in the House	"You can do this game at any point of the day, even a transition like when we are waiting for some children to finish breakfast" "Good for teaching addition and subtracting" "This activity taught me that children really have impressive memory and can add and subtract"	"Need more variation in the game. Children get bored after a while" "Different child skill levels makes this hard. Some children really don't get it. Some children really do" "Need more modeling on how to use the ten-frame"
Comparing Blocks	"Children were engaged by being able to participate and were engaged in comparing who has more" "Taught comparing of magnitudes well. I don't really have an activity that looks like this"	"Difficult to implement because have to come up with own materials" "Not very organized and is a too open-ended activity"
Dot Fun	"Helpful in learning the one-to-one correspondence skill I did not know before Teacher Lillie"	"Not very engaging. Children would benefit if the dots were animals or something else" "There are varying skill levels for a large group activity. If we keep this large group, have each child get their own dot cards to so can work at their own pace"

Shape Search Shape Sort	"Engaging for children" "Really noticed children making learning gains" "Like how they sorted color first then the size and shape. I really saw how the executive functions being used"	"Stronger lamination of shapes" "Intro to shapes first would be good. Some kids really don't know their shapes"
Construct Shape	"Children really learned difference between long and short" "Good at encouraging problem solving. Children have to try different ways with building a triangle in multi ways" "I don't really have activity that focuses on the sides of a shape like this. The children love the magnets and feels like building to them"	"Not enough materials for small group" "Again, intro to shapes would be helpful first"
Number Music	"Good to have movement and self-regulation with math. I like this freeze to the music concept in general"	"Sometimes hard to organize and some kids are misbehaving when waiting their turn" "Some children really are not recognizing number; how do they learn?"
Measuring Worms	"I liked how the book was related to the small group" "Children were engaged in this activity and different sized worms allowed them to problem solve how many would they need"	"Some kids still don't understand concept of measurement so intro to measurement would be helpful" "Some children just wanted to put the worms all around the edges of the animals, didn't understand to only measure one side" "How do we explain the difference between the length of worms, why don't we just use the 1-inch ones?"
Patterning	"Really helpful to know which skill comes first, then next with patterning skills"	"Some children cannot move on from the first level, how do I help them?"

	“Teacher Lillie made this activities levels super straight forward with the patterning templates- makes it easier to individualize during small-group” “Did not know about how to teach patterns and use A’s and B’s before this”	“Executive functioning level was too hard for children”
Pattern/Shape Blocks	“Children very engaged” “Can teach shape recognition but also fine motor skills and can ask children to count how many squares or something they used- teaches lots of skills in one activity”	“Some children it was hard for them to fit the blocks within the lines and they got frustrated”

Table IV.8. Emerging Themes of Teacher Knowledge Gains

Theme 1: Learned more math vocabulary

- “Learned more words for math like cardinality, one-to-one correspondence, trapezoid, rhombus”
- “Learned that children cannot grasp a quantity past 4”
- “Asking “how many?” is a key question and that this is different from assessing a child on if they count”
- “Didn’t know what a rhombus was”
- “That long and “short” are the correct words to use for teaching measurement I always use bigger and smaller, but I see that this is more for the teaching of sizes"

Theme 2: Learned more about children’s math skill readiness

- “Never knew that children could create shapes. Just thought we needed to teach what the shapes are called.”
- “Didn’t know children could extend patterns I always just teach them to repeat them”
- “I was not sure that children were ready for addition and subtraction in preschool"
- “We can use different objects from the classroom to measure things, and that kids actually are interested in measuring in this way. The measuring tape bored them but using the worms can make it fun”
- “Did not know that children could remember things so well in their brain, like with the working memory stuff we learned. How they could add and keep track of the monkeys was so crazy to me”

Table IV.9. Pairwise Correlations Between Study Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1) Experience	1.00																				
(2) Education	.02	1.00																			
(3) Und. math mid	.06	-.20	1.00																		
(4) Und. math post	.66*	.07	-.00	1.00																	
(5) Fidelity math	.33	.02	.28	.78*	1.00																
(6) Q's asked math	-.0	-.0	-.20	.67*	.63	1.00															
(7) Und. EF mid	-.0	.46	-.10	.30	.42	.40	1.00														
(8) Und. EF post	.20	.36	.13	.67*	.79*	.60	.20	1.00													
(9) Fidelity EF	-.0	-.1	.43	.35	.78*	.42	.11	.73*	1.00												
(10) Q's asked EF	.36	.19	.04	.66*	.78*	.56	.72*	.51	.33	1.00											
(11) Fidelity script	-.2	.37	.16	.17	.36	.11	.44	.37	.43	.13	1.00										
(12) Fidelity org	-.3	-.15	.42	.19	.69	.41	.27	.73*	.86*	.35	.68	1.00									
(13) Child engage.	.53	-.24	.26	.65*	.53	.21	-.2	.57	.31	.17	.29	.30	1.00								
(14) Attitude pre	.41	.01	.54	.70*	.84*	.45	.09	.59	.65	.54	.30	.56	.71*	1.00							
(15) Attitude post	.42	-.10	.37	.51	.56	.05	.00	.49	.28	.29	.54	.40	.86*	.62	1.00						
(16) Math conf. pre	.15	.78*	-.17	.35	.36	.11	.51	.17	.15	.42	.69*	.34	.19	.30	.34	1.00					
(17) Math conf. gains	-.1	-.73*	.30	-.10	-.1	-.0	-.4	.19	-.1	-.3	.02	.32	.49	.07	.44	-.48	1.00				
(18) Log compliance	-.2	-.3	.45	.10	.60	.23	-.1	.60	.89*	.05	.50	.86*	.38	.47	.44	-.03	.36	1.00			
(19) Math materials	-.5	-.5	.20	-.22	-.0	.15	-.3	.33	.43	-.4	.28	.62	.09	-.0	.00	-.39	.56	.64*	1.00		
(20) Math center	.35	-.0	.51	.60	.61	.37	-.0	.45	.59	.15	.30	.36	.63*	.81*	.47	.05	.15	.42	.27	1.00	
(21) Meet length	.41	.30	-.2	.78*	.56	.65*	.25	.48	.36	.39	.14	.20	.31	.49	.02	.40	-.41	-.07	-.09	.53	1.00
(22) Model length	.45	.39	-.3	.60	.23	.31	.50	.30	-.1	.34	.37	-.1	.17	.12	.10	.55	-.38	-.40	-.26	.20	.73*

*shows significance at .05 level

Study variables: teacher characteristics, fidelity of implementation, understanding of math/EF across activities, number of questions asked about math and EF across the intervention period, attitude towards intervention pre and post, confidence in teaching math pre and post, activity log compliance, and average length of coach meetings and coach modeling visits.

CHAPTER V: The Future of Math and Executive Functioning Preschool Interventions

The larger goal of this dissertation was to take a deep dive in preschool executive functioning and math interventions and uncover crucial factors to consider in the future design and evaluation of these approaches.

Understanding the "Black Box" in Preschool

This research first considered approaches to understanding the "black box" of EF and math interventions. In my analysis of the Tools of the Mind curriculum (Study 1), I focused first and foremost on the puzzling issue of poor fidelity of implementation in the Head Start CARES study and the startling increase in teacher use of scaffolding practices over time. Through this analysis, however, I discovered that a major obstacle with dissecting the "black box" was that data were limited to facets of the intervention and not to classroom practices or teaching beyond this context. Teachers may be practicing "EF" related support and scaffolding outside the confines of the targeted components, which in turn is impacting children's math skills. Weiss and colleagues stress the importance of collecting data on potential mediators of a treatment effect on treatment outcomes (Weiss, Bloom & Brock, 2014) but this has largely been neglected in evaluations of EF interventions. In the context of the Tools of the Mind intervention, my continuing hypothesis is that the consistent small impact on children's math skills was due to this application of knowledge to other areas of teaching, especially math, which relies heavily on scaffolding of skills from one math concept to the next.

Another issue in dissecting the “black box” of interventions was that few studies have collected data beyond fidelity of implementation. In Study III, I found that teachers’ understanding of the EF components and theory of change of the intervention was relatively weak, despite being able to implement the EF components to fidelity. This mismatch between teacher knowledge and observed practice has been highlighted at length with elementary school teachers (Akapame et al., 2019; Gess-Newsome et al., 2019; Konig & Pflanzl, 2016), but studies with early childhood educators have not typically considered both constructs. It is likely that with a less foundational understanding of EF skills, teachers’ fidelity to the EF component of the activities may decrease over time, and teachers may be less likely to use this knowledge in other teaching content areas. Yet in contrast to EF understanding, teachers in Study III did demonstrate a solid understanding of math development levels, but had more difficulty implementing to fidelity. It is possible that teachers may improve their fidelity with practice, as has been evidenced by the Building Blocks curriculum. Taken together, I have stressed the importance of not only observing what is happening on the outside (e.g., fidelity), but what is happening on the inside (e.g., teacher understanding), and how these two factors matter for understanding why an intervention was effective or not. Overall, future research aimed at understanding success of failure of an intervention would benefit from including teacher questionnaires that assess teachers’ understanding of an intervention as well as their application of practices to other areas of their teaching.

Combining Math and EF Content in Preschool

One of the major contributions of this research was also a demonstration of the promise of integrating math and EF content together (Studies II and III), and how this approach may be most beneficial for children of the lowest EF skills who also have at least a fundamental knowledge of cardinality. This contribution aligns with the recent influx of studies from early childhood researchers highlighting how poor EF skills can represent a significant barrier to children's math learning (DeFlorio et al., 2019; Dong et al., 2020; Ribner, 2020; Morgan et al., 2019; McCoy et al., 2019). This work also dovetails with efforts of prominent intervention researchers explicitly integrating EF content into existing math curricula (Clements et al., 2020; Mazzocco et al., in press) and math content into existing EF interventions (McClelland et al., 2019). Although the field has not reached consensus on the impact of this approach compared to math-only or EF-only efforts, increased pursuit of this combined content in the literature demonstrates the promise of this approach.

Importance of Educating Teachers on EF and Math Development

A final point I made in this dissertation is on the importance of educating preschool teachers on the development of EF and math skills. For EF specifically, policymakers have started to recognize EF as a crucial school readiness skill (National Institute for Early Education Research, Ackerman & Friedman-Krauss, 2017). However, state standard learning goals for EF in early childhood education settings lack clarity in detailing concrete goals for children's EF development and are often embedded within social-emotional skills or other cognitive domains such as motivation and creativity (California Preschool Learning Foundations, 2008; Department of Early Education and Care, 2015). This creates confusion for teachers struggling to support this important skill (Zelazo et al., 2017). Until the field can determine how to better educate teachers on EF skills, classroom interventions targeting this skill may continue to fall short.

Relatedly, it is also important to consider how to best educate teachers on EF development and the many ways it can manifest in the classroom beyond "paying attention." For example, the Head Start's Early Learning Outcomes Framework describes EF in the "cognitive self-regulation" category (2015) which includes "*child demonstrates an increasing ability to control impulses, child maintains focus and sustains attention with minimal adult support, child persists in tasks, child holds information in mind and manipulates it to perform tasks, child demonstrates flexibility in thinking and behavior."* Yet, again, there is little guidance for how to nurture these skills in context. There have been some efforts to align these "EF learning goals" with teacher practices by providing guidance to teachers on how to target these skills through CLASS domains (e.g., positive climate, teacher sensitivity, and behavior management; National Center on Early Childhood Development, Teaching, and Learning, 2018). However, these attempts to align EF skills with current teacher professional development materials and widely-used social-emotional curricula can create confusion for teachers struggling to distinguish EF from other foundational skills.

There are more resources, learning standards, and established curricula when it comes to improving teachers' understanding and teaching of math skills. Yet, preschool teachers still lack confidence in teaching math and STEM-related topics (Zinth, Weyer, & Atchison, 2019), and it is unclear whether this is due to little access to resources and curricula or a lack of knowledge regarding children's math development. Although a combination of the two is ideal support for a preschool teacher, teachers in Study III struggled most with practicing sensitivity to children's individual math skills when implementing the math/EF activities. Future work should continue to explore the best ways to educate teachers on math development, and identify barriers to motivating preschool programs to adapt evidenced-based math curricula that are thorough in providing a strong foundation of math development knowledge and curricular support (Clements & Sarama, 2008).

Concluding Thoughts

From a policy perspective, two key goals when evaluating childhood interventions are to identify which approaches are scalable, and which approaches maximize the impact on child outcomes. In this dissertation, I identified concrete barriers to achieving these two goals. For one, although I was able to educate teachers on children's math and EF development and teachers were able to implement combined content activities to fidelity, it was clear that teachers were more motivated to continue implementing an activity if they saw its direct connection to the CLASS and other tools they were already using. This preoccupation with improving CLASS quality scores presents an obstacle in coaching preschool teachers on anything other than what their preschool centers' funding depends on (e.g., QRIS scores), which also presents a major hindrance in achieving scalability with content-based interventions more generally. Similarly, for policymakers' second key goal of maximizing the impact of interventions on child outcomes, the field has largely neglected how *teachers* play a role in an intervention's effectiveness beyond teachers' fidelity of implementation, level of education, and teaching experience. *How* an intervention affects teachers' content-specific practice beyond the intervention components, *how* an intervention impacts teacher knowledge of the content, and *what* motivates teachers to continue using some or all parts of the intervention, are largely understudied and underemphasized in the literature. Just as policymakers want to know what interventions work for which children, I hope this dissertation sheds light on how they should also be asking the same question of teachers.

References

Ackerman, D.J., & Friedman-Krauss, A.H. (2017). Preschoolers' Executive Function: Importance, Contributors, Research Needs and Assessment Options. *ETS Research Report Series*, 1, 1-24.

Clements, D.H., Sarama, J., Layzer, C., Unlu, F., & Fesler, L. (2020). Effects on mathematics and executive function of a mathematics and play intervention versus mathematics alone. *Journal for Research in Mathematics Education*, *51*(3), 301-333.

DeFlorio, L., Klein, A., Starkey, P., Swank, P.R., Taylor, H.B., Halliday, S.E., Beliakoff, A., & Mulcahy, C. (2019). A study of the developing relations between self-regulation and mathematical knowledge in the context of an early math intervention. *Early Childhood Research Quarterly*, *46*(1), 33-48.

Dong, Y., Clements, D., Sarama, J., Dumas, D., Banse, H.W., & Day-Hess, C. (2020). Mathematics and Executive Function Competencies in the Context of Interventions: A Quantile Regression Analysis. *The Journal of Experimental Education*, 1-22.

Mazzocco, M.M., Chan, J., & Sera, M.D. (in press). Baseline EF skills are associated with children's responsiveness to different types of numeracy training.

Morgan, P.L., Farkas, G., Wang, Y., Hillemeier, M.M., Oh, Y., & Maczuga, S. (2019). Executive function deficits in kindergarten predict repeated academic difficulties in elementary school. *Early Childhood Research Quarterly*, *46*(1), 20-32.

McClelland, M.M., Tominey, S.L., Schmitt, S.A., Hatfield, B., Purpura, D., Gonzales, C., &

Tracy, A. (2019) Red Light, Purple Light! Results of an intervention to promote school readiness for children from low-income backgrounds. *Frontiers in Psychology, 10*.

McCoy, D.C., Gonzalez, K., & Jones, S. (2019). Preschool Self-regulation and Preacademic Skills as Mediators of the Long-Term Impacts of an Early Intervention. *Child Development*, *90*(5), 1544-1558.

Ribner, A.D. (2020). Executive function facilitates learning from math instruction in kindergarten: Evidence from the ECLS-K. *Learning and Instruction*, 65.

Zelazo, P.D., Blair, C.B., & Willoughby, M.T. (2017). Executive function: Implications for Education. NCER 2017-2000. *National Center for Education Research*

Zinth, J., Weyer, M., Atchison, B. (2019). Enhancing STEM in P-3 Education. Policy Guide. Education Commission of the States

Appendix A: Full Results Tables for Study 1

Appendix A Table 1. Child Baseline Differences Between Tools (*N*=457) and Control (*N*=448) Group

	Control	Tools	Difference	Effect size
Age (years)	4.40	4.42	.02	.03
White	14.84	19.24	4.40+	.10
Black	26.31	26.44	.13	.05
Latinx	48.88	47.18	-1.70	.04
Other	9.78	7.47	-2.32	-.08
% Female	49.52	48.86	-.66	-.03
Head-to-Toes	2.41	2.50	.09	.01
Pencil Tap	.45	.46	.01	.01
WJ Applied Problems	395.36	396.60	1.24	.07

F (9,896) = 133.73, p=.39

Note. ***p<.001 **p<.01 *p<.05 +p<.10

Appendix A Table 2. Teacher Baseline Differences Between Tools (*N*=76) and Control (*N*=77) Group

	Control	Tools	Difference	Effect size
Age (years)	40.38	43.83	3.45	.10
White	28.02	27.25	-.77	.08
Black	30.68	30.40	-.27	-.02
Latinx	36.24	34.38	-1.86	-.04
Other	4.54	7.95	3.41	.07
% Have Bachelor's degree or higher	56.92	56.81	.11	.02
% Have years of teaching experience > 10	68.07	74.24	6.17*	.11

F (7,146) = 127.04, p=.20

Note. ***p<.001 **p<.01 *p<.05 +p<.10

Appendix A Table 3. Differences at Baseline Between Children Who Stayed and Children Who Left the Study

	Stayed (*N*=1,114)	Left (*N*=185)	Difference	Effect size
Age (years)	4.40	4.33	-.07	-.08
White	17.08	18.52	1.44	.05
Black	26.03	25.89	-.14	-.01
Latinx	48.01	46.09	-1.92	-.04
Other	6.59	8.18	1.59	.07
% Female	48.14	42.19	-5.95**	-.13
Head-to-Toes	2.26	2.02	-.24	-.05
Pencil Tap	.46	.41	-.05+	-.14
WJ Applied Problems	394.31	391.45	-2.86+	-.10

F (9,1290) = 189.56, p=.38

Note. ***p<.001 **p<.01 *p<.05 +p<.10

Appendix A Table 4. Replication Analysis of Tools' Impact on EF and Math Skills

	Math	Head-to-Toes	Pencil Tap
INT: Tools	2.79(1.64)+	-.04(.24)	-.00(.02)
Baseline	.57 (.02)***	.43(.03)***	.37(.02)***
Baseline imputed flag	1.19 (2.10)	-.37(.44)	-.11(.04)***
INT: Incredible Years	.42(1.63)	-.16(.24)	.00(.02)
INT: PATHS	2.16(1.65)	-.24(.24)	.02(.02)
Expressive vocab	.28(.04)***	.05(.006)***	.00(.00)**
Expressive vocab imp	-4.53(2.11)***	.27(.44)	.05(.04)
Child age	1.22(.13)***	.22(.02)***	.01(.00)
Block ID= 2	-.96(4.83)	1.25(.71)+	.01(.06)
Block ID= 3	-12.92(3.11)***	.70(.43)	-.00(.03)
Block ID= 4	-10.97(3.33)***	-1.08(.46)*	-.02(.04)
Block ID= 5	-.44(3.21)	.53(.45)	-.07(.04)+
Block ID= 6	-4.70(2.63)*	1.21(.37)**	-.02(.03)
Block ID= 7	-5.97(3.82)	.44(.53)	-.11(.04)*
Block ID= 8	-5.43(3.12)*	-.83(.44)+	.01(.04)
Block ID= 9	-10.76(3.38)***	-.91(.47)+	-.06(.04)
Block ID= 10	-9.94(4.23)**	.43(.61)	.03(.05)
Block ID= 11	-12.74(3.75)***	-.93(.52)+	-.13(.04)***
Block ID= 12	-5.38(4.22)	.78(.59)	.06(.05)
Block ID= 13	2.72(3.29)	.80(.46)+	.03(.04)
Block ID= 14	-5.15(3.82)	.10(.54)	-.07(.04)+
Block ID= 15	-5.95(2.47)*	.12(.34)	-.06(.03)
Block ID= 16	-9.61(4.57)**	.22(.67)	-.11(.05)*
Block ID= 17	-14.77(3.74)***	-.62(.54)	-.13(.04)**
Block ID= 18	-6.94(3.36)**	-.32(.45)	.04(.04)
Block ID= 19	-11.84(3.48)***	.18(.47)	-.00(.04)
Block ID= 20	-3.59(3.28)	.55(.46)	.02(.04)
Block ID= 21	-8.25(2.95)***	.10(.42)	-.08(.03)*
Block ID= 22	-4.40(3.35)	.84(.46)+	-.02(.04)
Constant	107.10(9.92)***	-13.37(1.30)***	-.36(.10)***
Children	2,584	2,596	2,599
Classrooms	305	305	305

Note. Each model controls for child age, baseline WAP score, baseline expressive language score, and a flag for whether the baseline score was imputed or not. Random intercepts are at the center and classroom level and there is a fixed effect for block. This model design is identical to the one used in the original Head Start CARES analyses. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 5. Fidelity of Implementation: Play Planning

	Math	Head-to-Toes	Pencil Tap
Average Play Planning	-.30(1.30)	.21(.04)	.01(.01)
Baseline	.57(.04)***	.49(.05)***	.31(.04)***
Baseline imputed flag	.31(4.60)	-.97(1.05)	-.06(.08
Expressive vocab	.18(.07)	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-5.63(4.62)	.88(1.04)	-.01(.08)
Child age	1.42(.26)***	.26(.04)***	.02(.00)***
Initial Play Planning	3.97(1.13)**	.18(.18)	.00(.01)
Baseline	.56(.04)***	.50(.06)***	.31(.05)***
Baseline imputed flag	1.11(5.11)	.16(1.27)	.01(.10)
Expressive vocab	.11(.08)	.04(.01)***	.00(.00)**
Expressive vocab imputed flag	-8.11(5.14)	-.19(1.27)	-.10(10)
Child age	1.64(.30)***	.22(.05)***	.02(.00)***
End Play Planning	-.07(1.0)	-.15(.16)	-.03(.01)*
Baseline	.57(.04)***	.49 (.05)***	.30(.04)***
Baseline imputed flag	-.77(4.70)	-1.38(1.09)	-.06(.08)
Expressive vocab	.19(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-5.70(4.74)	1.22(1.09)	-.01(.08)
Child age	1.36(.27)***	.24(.04)	.02(.00)***

Note. For math outcomes N=656 children. For Head-to-Toes and Pencil Tap outcomes, N=658 children. For all outcomes N=75 centers. Each model controls for child age, baseline WAP score, baseline expressive vocabulary, and flags for whether the baseline score was imputed or not. Random intercepts at the center level. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Plan = Plan planning. Initial fidelity was calculated averaging September and October scores, and End fidelity was calculated averaging March and April scores. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 6. Fidelity of Implementation: Make-Believe Play

	Math	Head-to-Toes	Pencil Tap
Average Make-Believe Play	-.85(1.17)	.15(.19)	.00(.01)
Baseline	.57(.04)***	.50(.05)***	.31(.04)***
Baseline imputed flag	.48(4.59)	-.98(1.05)	-.06(.08)
Expressive vocab	.18(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-5.69(4.62)	.89(1.04)	-.01(.08)
Child age	1.43(.26)***	.26(.04)***	.02(.00)***
Initial Make-Believe Play	-.02(1.32)	.24(.19)	.01(.01)
Baseline	.57(.04)***	.50(.06)***	.30(.05)***
Baseline imputed flag	1.72(5.08)	.13(.24)	.01(.10)
Expressive vocab	.16(.08)	.04(.01)***	.00(.00)
Expressive vocab imputed flag	-7.96(5.11)	-.18(1.27)	-.09(.10)
Child age	1.69(.30)***	.22(.04)***	.02(.00)***
End Make-Believe Play	-.90(.90)	.22(.14)	.01(.01)
Baseline	.57(.04)***	.50(.05)***	.30(.04)***
Baseline imputed flag	-.59(4.69)	-1.46(1.09)	-.06(.08)
Expressive vocab	.19(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-5.75(4.74)	1.28(1.09)	-.01(.08)
Child age	1.36(.37)***	.24(.04)***	.02(.00)***

Note. For math outcomes N=656 children. For Head-to-Toes and Pencil Tap outcomes, N=658 children. For all outcomes N=75 centers. Each model controls for child age, baseline WAP score, baseline expressive vocabulary, and flags for whether the baseline score was imputed or not. Random intercepts at the center level. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Plan = Plan planning. Initial fidelity was calculated averaging September and October scores, and End fidelity was calculated averaging March and April scores. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 7. Fidelity of Implementation: Make-Believe Play Practice

	Math	Head-to-Toes	Pencil Tap
Average Make-Believe Play Practice	.44(1.14)	.18(.20)	.00(.01)
Baseline	.57(.04)***	.50(.05)***	.31(.04)***
Baseline imputed flag	.58(4.62)	-.94(1.04)	-.06(.08)
Expressive vocab	.16(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-6.04(4.67)	.89(1.04)	-.01(.08)
Child age	1.40(.26)***	.26(.04)***	.02(.00)***
Initial Make-Believe Play Practice	2.52(1.62)	.31(.24)	.01(.01)
Baseline	.56(.04)***	.50(.06)***	.30(.05)***
Baseline imputed flag	1.66(5.07)	.16(1.27)	.01(.10)
Expressive vocab	.15(.08)	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-8.19(5.11)	-.21(1.27)	-.09(.10)
Child age	1.68(.30)***	.22(.05)***	.02(.00)***
End Make-Believe Play Practice	.94(.79)	.13(.14)	.01(.01)
Baseline	.59(.04)***	.49(.05)***	.30(.04)***
Baseline imputed flag	-.51(4.73)	-1.39(1.09)	-.06(.08)
Expressive vocab	.17(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-6.03(4.78)	1.25(1.09)	-.01(.08)
Child age	1.34(.27)***	.24(.04)***	.02(.00)***

Note. For math outcomes N=656 children. For Head-to-Toes and Pencil Tap outcomes, N=658 children. For all outcomes N=75 centers. Each model controls for child age, baseline WAP score, baseline expressive vocabulary, and flags for whether the baseline score was imputed or not. Random intercepts at the center level. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Plan = Plan planning. Initial fidelity was calculated averaging September and October scores, and End fidelity was calculated averaging March and April scores. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 8. Fidelity of Implementation: Buddy Reading

	Math	Head-to-Toes	Pencil Tap
Average Buddy Reading	-.89(1.28)	.15(.21)	.01(.01)
Baseline	.57(.04)***	.50(.05)***	.30(.04)***
Baseline imputed flag	.53(4.59)	-.94(1.05)	-.06(.08)
Expressive vocab	.18(.07)*	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-5.72(4.62)	.87(1.04)	-.01(.08)
Child age	1.43(.26)***	.26(.04)***	.02(.00)***
Initial Buddy Reading	1.78(1.41)	.16(.21)	.01(.01)
Baseline	.55(.04)***	.50(.06)***	.30(.05)***
Baseline imputed flag	1.49(5.08)	.19(1.27)	.02(.10)
Expressive vocab	.16(.08)	.04(.01)***	.00(.00)***
Expressive vocab imputed flag	-7.95(5.11)	-.22(1.27)	-.10(.10)
Child age	1.67(.30)***	.22(.05)	.02(.00)***
End Buddy Reading	-.83(1.01)	.11(.17)	.01(.01)
Baseline	.57(.04)***	.49(.05)	.30(.04)***
Baseline imputed flag	-.61(4.70)	-1.38(1.09)	-.06(.08)
Expressive vocab	.19(.07)*	.04(.01)***	.00(.00)
Expressive vocab imputed flag	-5.78(4.74)	1.23(1.09)	-.01(.08)
Child age	1.37(.27)***	.24(.04)***	.02(.00)***

Note. For math outcomes N=656 children. For Head-to-Toes and Pencil Tap outcomes, N=658 children. For all outcomes N=75 centers. Each model controls for child age, baseline WAP score, baseline expressive vocabulary, and flags for whether the baseline score was imputed or not. Random intercepts at the center level. BR= Buddy reading, MBP= Make-believe play, MBPP= Make-believe play practice, Plan = Plan planning. Initial fidelity was calculated averaging September and October scores, and End fidelity was calculated averaging March and April scores. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 9. Sample Restricted to Teachers with a BA Predicting Math and EF Skills

	Math	Head-to-Toes	Pencil Tap
INT: Tools	.83(2.35)	.19(.33)	.01(.03)
Baseline	.57(.03)***	.44(.04)***	.36(.03)***
Baseline imputed flag	-2.45(3.19)	-.92(.62)	-.10(.05)+
INT: Incredible Years	-1.07(2.31)	.37(.33)	.01(.03)
INT: PATHS	.27(2.31)	-.46(.33)	.02(.03)
Expressive vocab	.28(.05)***	.06(.01)***	.00(.00)***
Expressive vocab imp	.03(3.20)	.97(.63)	.04(.05)
Child age	1.08(.19)***	.20(.03)***	.01(.00)***
Block ID= 2	-8.52(10.40)	-.03(1.56)	-.04(.14)
Block ID= 3	-7.02(8.53)	2.24(1.15)+	.01(.10)
Block ID= 4	-11.86(4.58)***	-1.80(.65)***	.00(.05)
Block ID= 5	-.08(4.19)	.45(.59)	-.09(.05)+
Block ID= 6	-.08(4.19)	.10(.47)*	-.01(.04)
Block ID= 7	-9.32(5.06)	.08(.72)	-.11(.06)+
Block ID= 8	-5.87(4.74)	-.73(.66)	.03(.06)
Block ID= 9	-11.98(3.85)***	-1.19(.53)*	-.07(.04)
Block ID= 10	-14.12(6.66)**	-.48(.98)	.04(.08)
Block ID= 11	-9.71(5.84)*	-2.72(.84)***	-.18(.07
Block ID= 12	--	---	---
Block ID= 13	1.86(4.20)	.75(.60)	.00(.05)
Block ID= 14	-8.26(4.29)	-.38(.61)	-.08(.05)
Block ID= 15	-5.13(3.21)	-.26(.46)	-.06(.04)
Block ID= 16	-18.94(7.67)*	.37(1.15)	-.03(.09)
Block ID= 17	-21.22(6.87)***	-.14(1.03)	-.22(.08)***
Block ID= 18	-3.26(4.21)	-.71(.57)	.01(.05)
Block ID= 19	-11.29(4.86)*	-.90(.65)	-.04(.06)
Block ID= 20	-.24(4.74)	.28(.68)	.07(.06)
Block ID= 21	-15.56(4.93)***	.19(.72)	-.07(.06)
Block ID= 22	-5.35(3.59)	.71(.50)	-.02(.04)
Children	1,240	1,250	1,247
Classrooms	134	134	134

Note. Tools' impact on math skills restricted to Teachers with a BA. Random intercepts are at the classroom and center level and there is a fixed effect for Block ID. +p<.10 *p<.05 ** p<.01 ***p<.001

Appendix A Table 10. Sample Restricted to Teachers Without a BA

	Math	Head-to-Toes	Pencil Tap
INT: Tools	3.66(2.56)	.03(.13)	-.01(.03)
Baseline	.62(.04)***	.43(.05)***	.36(.04)***
Baseline imputed flag	2.48(4.11)	-.36(.84)	-.14(.07)+
INT: Incredible Years	2.79(2.87)	-.64(.47)	.01(.04)
INT: PATHS	4.04(2.81)	.04(.46)	.04(.03)
Expressive vocab	.30(.07)***	.06(.01)***	.00(.00)***
Expressive vocab imp	-9.31(4.13)*	.18(.84)	.08(.07)
Child age	1.10(.25)***	.19(.04)***	.01(.00)***
Block ID= 2	7.00(7.86)	.41(1.30)	-.04(.10)
Block ID= 3	-10.57(6.32)	-.59(1.06)	-.04(.08)
Block ID= 4	-8.11(8.18)	-2.15(1.36)	-.08(.10)
Block ID= 5	7.97(8.21)	-.32(1.33)	-.04(.10)
Block ID= 6	1.33(6.69)	.11(1.10)	-.10(.08)
Block ID= 7	2.25(7.29)	-.48(1.18)	-.16(.09)+
Block ID= 8	1.46(6.43)	-1.64(1.08)	-.05(.08)
Block ID= 9	---	---	---
Block ID= 10	-9.63(8.23)	-1.07(1.36)	-.09(.10)
Block ID= 11	12.73(9.75)	-1.66(1.61)	-.04(.12
Block ID= 12	-1.54(8.13)	-.15(1.30)	-.04(.12)
Block ID= 13	4.63(9.46)	-2.04(1.57)	.08(.12)
Block ID= 14	---	---	---
Block ID= 15	-1.99(6.23)	-.30(1.02)	.07(.08)
Block ID= 16	.85(8.27)	-.79(1.37)	-.25(.10)*
Block ID= 17	-16.85(7.65)*	-2.68(1.26)*	-.21(.09)*
Block ID= 18	---	---	---
Block ID= 19	-3.93(6.76)	.32(1.13)	-.02(.08)
Block ID= 20	-7.76(7.44)	-.26(1.21)	-.08(.09)
Block ID= 21	-1.48(6.50)	-.73(1.08)	-.13(.08)
Block ID= 22	---	---	---
Constant	84.08(19.32)***	-11.41(2.69)***	-.44(.20)***
Children	697	697	702
Classrooms	88	58	88

Note. Tools' impact on math skills restricted to Teachers without a BA. Random intercepts are at the classroom and center level and there is a fixed effect for Block ID. +*p*<.10 **p*<.05 ** *p*<.01 ****p*<.001

Appendix A Table 11. Combined Model with Interaction Between Teacher Education and Treatment

	Math	Head-to-Toes	Pencil Tap
Teacher Ed*Tools	2.16(1.49)+	-.14(.52)	.02(.04)
Baseline	.62(.04)***	.43(.05)***	.36(.04)***
Baseline imputed flag	2.48(4.11)	-.36(.84)	-.14(.07)+
Expressive vocab	.30(.07)***	.06(.01)***	.00(.00)***
Expressive vocab imp	-9.31(4.13)*	.18(.84)	.08(.07)
Child age	1.10(.25)***	.19(.04)***	.01(.00)***
Block ID= 2	7.00(7.86)	.41(1.30)	-.04(.10)
Block ID= 3	-10.57(6.32)	-.59(1.06)	-.04(.08)
Block ID= 4	-8.11(8.18)	-2.15(1.36)	-.08(.10)
Block ID= 5	7.97(8.21)	-.32(1.33)	-.04(.10)
Block ID= 6	1.33(6.69)	.11(1.10)	-.10(.08)
Block ID= 7	2.25(7.29)	-.48(1.18)	-.16(.09)+
Block ID= 8	1.46(6.43)	-1.64(1.08)	-.05(.08)
Block ID= 9	---	---	---
Block ID= 10	-9.63(8.23)	-1.07(1.36)	-.09(.10)
Block ID= 11	12.73(9.75)	-1.66(1.61)	-.04(.12
Block ID= 12	-1.54(8.13)	-.15(1.30)	-.04(.12)
Block ID= 13	4.63(9.46)	-2.04(1.57)	.08(.12)
Block ID= 14	---	---	---
Block ID= 15	-1.99(6.23)	-.30(1.02)	.07(.08)
Block ID= 16	.85(8.27)	-.79(1.37)	-.25(.10)*
Block ID= 17	-16.85(7.65)*	-2.68(1.26)*	-.21(.09)*
Block ID= 18	---	---	---
Block ID= 19	-3.93(6.76)	.32(1.13)	-.02(.08)
Block ID= 20	-7.76(7.44)	-.26(1.21)	-.08(.09)
Block ID= 21	-1.48(6.50)	-.73(1.08)	-.13(.08)
Block ID= 22	---	---	---
Constant	84.08(19.32)***	-11.41(2.69)***	-.44(.20)***
Children	697	697	702
Classrooms	88	58	88

Note. Unrestricted model with interaction between Teacher Education (0/1) and Treatment (0/1). Random intercepts are at the classroom and center level and there is a fixed effect for Block ID.
+p<.10 *p<.05 ** p<.01 ***p<.001

Appendix B: Study III Measures

Appendix B.I

Math Observation (October 2019)

Lead Teacher:	Date:
# of Assistant Teachers:	Start/End Time:
Number of children:	Half-day/Full-day

Math Center: yes/no Materials out: yes/no Number of children attended: ___________ Teacher at center: yes/no Teacher providing instructional support: yes/no	Describe:
Math Small group: yes/no Materials used:__________________ Target math skill: ________________ Number of children per table:	Describe:
Math Large group: yes/no Materials used:__________________ Target math skill:_________________	Describe:
Math Meal time/Brush teeth: yes/no	Describe:
Math talk during transitions: yes/no	Describe:
Observed math errors	
Materials out in the classroom (visible on shelves)	
Materials available in storage (ask teacher)	

Appendix B.II
Coach Survey (September 2019)

<u>Executive Functioning Understanding</u>

(1) Please articulate how children exercise their executive functioning in the below activities:

1a. Freeze Game
1b. Number Search
1c. Monkeys in the House

(2) Please articulate why you think strengthening children's executive functioning is important?

<u>Math Skill Development Understanding: Operations, Geometry, Number Sense, Patterning</u>

(3) For the Monkeys in the House, where should the teacher start with a child of the lowest skill level?

3a. Once a child masters this skill, what should the teacher scaffold for them next?

3b. What is the "hardest" level in this game/the most advanced skill?

(4) For the Shape Search Activity, what should a child of the lowest skill level be asked to do first?

4a. Once a child masters this skill, what should the teacher scaffold for them next?

4b. What is the "hardest" level in this game/the most advanced skill?

(5) For the Dot Fun activity, what card should be shown to the child of the lowest skill level to count and compare?

5a. Once a child masters this skill, what should the teacher scaffold for them next?

5b. What is the "hardest" level in this game/the most advanced skill?

(6) For the Patterning activity, what should a child of the lowest skill level be asked to do first?

6a. Once a child masters this skill, what should the teacher scaffold for them next?

6b. What is the "hardest" level in this game/the most advanced skill?

(7) Are you anticipating any struggles with coaching the teachers? Describe:

Appendix B.III
Teacher Survey- Pre (October 2019)

(1) How confident are you in teaching math? (circle one)

1= not confident   2=somewhat confident   3= confident   4=very confident

(2) What Curriculum(s) do you use?

(3) Do you use this Curriculum for math activity ideas, or do you use another resource for math instruction? Please explain.

(4) Have you ever received math coaching before? Check all that apply:
____ I have only attended a workshop or training where I learned how to teach math without being coached
____ I have received training on how to increase my use of math language in the classroom
____ I have been provided with specific math activities to implement before
____ I have been provided with modeling and observation of how to improve my math instruction before
____Other (please elaborate):

(5) How long have you been teaching preschool? _________
(6) How long have you been teaching Head Start preschool? _______________
(7) How long have you been teaching at your current center? ______________
(8) What is your age? _________________
(9) What is your highest level of education?
____Associates
____Bachelors
____Masters
Other:____________________

(10) Spanish use in the classroom (Check all that apply)

____ I do not use Spanish
____ I only use Spanish with specific students who know very little English
____ I use Spanish occasionally during songs, or when learning new vocab words—I do not use Spanish to give directions or instruction
____ I switch between English and Spanish when teaching
----
(11) For Lillie to rate at time of Pre-Coaching Survey: Rate teachers' overall interest and attitude in this intervention

1= not interested   2=mildly interested   3=very interested

Appendix B.IV
Teacher Survey- Mid (December 2019)

<u>Executive Functioning Understanding</u>

(1) Please articulate how children exercise their executive functioning in the below activities:

1a. Freeze Game

1b. Number Search

1c. Monkeys in the House

(3) Please articulate why you think strengthening children's executive functioning is important?

<u>Math Skill Development Understanding: Operations, Geometry, Number Sense</u>

(4) For the Monkey's in the House activity, what should you do with a child of the lowest skill level?

4a. Once a child masters this skill, what should you scaffold for them next?

4b. What is the "hardest" level in this game/the most advanced skill?

(5) For the Shape Search activity, what should a child of the lowest skill level be asked to do first?

5a. Once a child masters this skill, what should you scaffold for them next?

5b. What is the "hardest" level in this game/the most advanced skill?

(6) For the Dot Fun activity, what card should be shown to the child of the lowest skill level to count and compare?

6a. Once a child masters this skill, what should the teacher scaffold for them next?

6b. What is the "hardest" level in this game/the most advanced skill?

(7) Rate your confidence in implementing the activities on your own:

1= not confident 2= somewhat confident 3= confident 4=very confident

Appendix B.V

Teacher Survey- Post (March 2020)

Confidence & Understanding of Intervention

(1) Would you say your confidence has changed in teaching math since participating in the coaching? (circle one)

No- it hasn't changed Yes; I am *a little* more confident Yes; I am *a lot* more confident

(2) Please rate your top 3 favorite activities:

1________________ Why:

2________________ Why:

3________________ Why:

Executive Functioning Understanding

(3) Please articulate how children exercise their executive functioning in the below activities:

3a. Freeze Game

3b. Shape Search and Sort

3c. Monkeys in the House

3d. Measuring Worms

(Etc. Repeat for Remaining Activities)

(4) Please articulate why you think strengthening children's executive functioning is important?

Math Skill Development Understanding:

(5) For the Number Search activity, what should you do with a child of the lowest skill level?

5a. Once a child masters this skill, what should you scaffold for them next?

5b. What is the "hardest" level in this game/the most advanced skill?

(6) For the Monkeys in the House activity…

(7) For the Patterning Templates activity…

(Etc. Repeat for Remaining Activities)

<u>Rating of Your Coaching Experience</u>

(10) Please rate how well you were coached on the below activities from Lillie:

10a. Number Search activity
1= poor 2= ok 3= good 4= excellent N/A: Was not coached by Lillie on this activity

Etc. (repeat for all 12 activities)

(11) Please rate how well you were coached on the below activities from your Coach:

11a. Number Search activity
1= poor 2= ok 3= good 4= excellent N/A: Was not coached by Heidi on this activity

Etc. (repeat for all 12 activities)

(12) Please indicate your preferred frequency of coaching:

__ Weekly (like you received this year)
__ Bi-weekly (every other week)
__ Monthly
__ Few times per year
__ Other:

(13) Please indicate how you like to receive coaching (choose best option that applies):

___ Modeling and then observation and feedback (like you received this year) in the classroom
___ Modeling-only in the classroom; do not want follow-up/feedback
___ Modeling *outside* of the classroom
___ Just provide materials and the activity booklet- no modeling

<u>COVID-19 Addition</u>

14) Describe your daily life in quarantine/how you spend your days
15) How would you compare your stress level to before quarantine?
16) What are you most concerned about for the children in your classroom?
17) What are you most concerned about for your future work as a teacher?

For Lillie to rate at the time this Post-Coaching Survey is administered:
Rate teachers' overall interest and attitude in this intervention (post)

1= not interested 2=mildly interested 3=very interested

Appendix B.VI
Teacher Activity Log(s) (January – March 2020)

Date ___________

Which activities did you implement today from the math/EF activity booklet?

1.) ____________________ (insert activity name)

______ (insert level(s) you implemented within this activity)

How long did this activity last? ___________________

On a scale of 1-4, how engaged were children? __________________

(*1= Most children are not-engaged and either bored or restless, 2= Around half of the children are engaged, but the other half are acting bored or restless, 3= Most children are engaged except a few children, 4= All children are engaged)*

What went well with this activity?

What was difficult?

<u>Documented children learning gains:</u>

Child Name:________________

What did this Child learn today/did they get through a level?

Child Name:________________

What did this Child learn today/did they get through a level?

**Teacher fills out a log for each activity they did; Can add as many child names as they want, but have to write down at least 2 names each time they do an activity*

Appendix B.VII

Fidelity of Implementation (March 2020)

Date: Start/End Time:
Lead teacher: Observer:
Assessment teachers: Large or small group
Children present:

Adherence 1= went completely off-script 2= stayed on script but added in own elements 3= stayed on script
Child Engagement 1= most children not engaged 2= around half of the children are engaged 3= most children are engaged
Quality *EF integration* 1= did not do it at all 2= only executed part of it 3= appropriately executed str *Math skill level* 1= started at a too difficult level 2= started at correct level but did not move on when majority children demonstrated readiness; or moved on too quickly 3=started at correct level and moved on when majority children demonstrated readiness *Organization/Management* 1= materials not set up ahead of time and/or most children off-task 2= teacher is organized with materials but is only focusing on one or two children for too long/other children off-task 3=teacher is organized with materials and most children included
Dosage 1= activity was too short/children did not have an opportunity to try another level; or only 1 child received a turn at large group 2= teacher went through appropriate levels, but activity was too long/children began to grow restless 3= activity length was just right

Overall Fidelity is calculated by taking the average across all categories (Range 1-3).

Double Coded By: _______________ and _________________

Appendix B.VIII. Fidelity of Implementation Rating Rubric

	Organization	EF Integration	Math Sequence	Child Engagement	Adherence to script
1= poor	Materials not set up prior, or teacher was only focusing on one or two children and other children were off-task (small-group) or teacher taking too long with one child's turn (at large group)	Did not happen	Started at level that was too hard for child (if small group activity), or majority of children (if large group activity)	Most children are not engaged	Teacher completely changed activity
2= fine	A little chaotic with a few children off-task, but overall OK and not a lot of time was wasted	Happened, but only partially	Started at level that was too easy for child(ren), or it was the correct level, but the teacher did not scaffold to the next level	Around half the class is engaged, but the rest are not engaged	Teacher followed some elements of the script, but added other storylines or rules into the activity
3= spot on	Went smoothly from start to finish	Happened fully	Started at appropriate level for child and scaffolding to move them to next level	Most children are engaged	Teacher followed script

Appendix B.IX
Teacher Interview (April 2020)

<u>Content</u>

(1) What were the strengths and weaknesses of each of the activities?

(Go through each activity)

(2) Did you learn anything new about how to teach addition and subtraction?

(3) Did you learn anything new about how to teach patterning skills?

(4) Did you learn anything new about how to teach counting, and understanding of quantity?

(5) Did you learn anything new about how to teach number recognition?

(6) Did you learn anything new about how to teach shape recognition and shape attributes?

(7) Did you learn anything new about how to teach measurement?

(8) What did you learn about how executive functioning was incorporated in each activity? Do you think this integration was helpful?

<u>Coaching</u>

(9) What did you enjoy about the coaching this year?

(10) Were there any barriers to receiving math coaching this year? What could have made the learning of the activities *easier* for you?

(11) How would you improve the math coaching next year/How would you like to receive math coaching in the future?

Appendix B.X
Coach Interview (April 2020)

Content

(1) What were some strengths of the activities themselves?

(2) What were some weaknesses?

Coaching

(3) What worked about this coaching framework?

(4) What didn’t work?

(5) How would you improve the coaching moving forward?

Appendix B.XI.

Teacher and Coach Articulation of EF Integration and Math Level Understanding Scoring Rubric (Teacher and Coach Survey)

Score	EF Integration	Math level
1	Uses EF-related language, but does not connect this to any specific facet of the activity Example: "Children really have to pay attention to the teacher and what she is doing"	Can identify one level correctly (even if vague in their description) but cannot identify any more levels than that Example: "The first level is just being able to count the bears, then child has to keep track of bears in the house"
2	Uses EF-related language and connects it to a specific facet of the activity, but it is the incorrect facet OR identifies the correct facet of the activity that incorporates EF, but describes the type of EF used incorrectly Example: "Children have to switch attention between the houses to keep track of the numbers"	Can identify two levels, but is either incorrect with identifying the rest, or the levels are in the incorrect sequence, or fails to describe any more levels before or after (if there are more than two). Example: "Children have to add 1 bear, and then subtract 1 bear, and then add 2 bears, subtract 2 bears"
3	Connects the specific aspect of the activity that integrates EF, and uses the correct terminology when describing this connection- But description is vague Example: "Children use their memory to remember how many monkeys went into the house"	Can identify the general gist of the sequence of skills within the activity, but is vague in their description Example: "The first level is just saying how many bears are in a house, then the child has to keep track of adding bears, subtracting bears, and then you can add two houses if child is ready"
4	Identifies everything for a scoring of 3, but describes the integration more in detail Example: "Children use working memory when they keep track of how many monkeys are added or subtracted from the house without seeing the monkeys. They have to update this number in their head"	Can identify the sequence of math skills within the activity and describes in detail Example: "The first level is understanding "how many" bears are in a house (no more than 4 bears in the house at all times), then the child has to keep track of adding bears (1 bear, then 2 bears), subtracting bears (1 bear, tops). For some children, you can use the ten-frame to scaffold this until their memory is improved. Then you repeat the process, but have two houses so that children's working memory is even more strengthened while learning adding and subtracting"

Appendix C: Results from the Baseline Classroom Math Observation Measure

Appendix C: Results from the Baseline Classroom Math Observation Measure

Classroom materials. There was wide variability in the number and type of materials available in classrooms (see Table IV.2). Four teachers had an "abundant" number of math materials including geometry materials, patterning, printed numerals, and counters. Four teachers had "adequate" materials that included the above *variety* but did not have enough *quantity* for a small group activity (e.g., where each child was able to work with materials). Two teachers were "lacking" in their materials as they did not have any counters and only had some shape blocks and number recognition cards.

Observed math instruction. A summary of observed classroom math practices is in Table IV.3.

Math center. Similar to materials, there was also variability in the use of the math center in each of the classrooms. For example, all classrooms had a math center where children could choose to go to and engage with math-related materials. However, during center time, I identified three different levels of quality from my observational data. The lowest level of quality was observed in five classrooms where there were no materials/suggested activities out on the table (but there were math materials available in the shelves). In these "lowest level" classrooms, on average, two children visited the math center during a 60-minute center-time block. The second level of quality was observed in three classrooms where the center had specific materials out on the table (e.g., shape pattern blocks and templates, Magna tiles, numeral cards with counter people) but teachers did not provide any instructional support on how to interact with the materials (at least that I could observe). In these classrooms, on average, five children visited this center during the 60-minute center-time block. The highest level of quality was observed in two classrooms where teachers put materials out on the table and then provided instructional support during center-time to children attending the math center (e.g., teacher showed children how to pattern with beads, teacher led a card game matching numbers withs dots). In these classrooms, on average, seven children visited this center during the 60-minute center-time block.

Small group. For four out of the 10 teachers, I also observed a small-group math activity. For these activities, I categorized quality into two groups. For the first category (three teachers) the activity only involved passing out materials and having children complete a task without any explicit instruction. For example, for one observation the teacher told children to cut out numerals and match them to the quantity of dots, and then glue them together on a piece of paper (this was the only direction). Throughout the activity the teacher was observed sometimes scaffolding one-to-one correspondence for a child by holding their finger one-to-one on each dot, but overall this was a chaotic activity that ended up focusing more on cutting out the numbers and gluing them to the piece of paper than focusing on numeral recognition and quantity. In fact, more than half of the children had difficulty with numeral-quantity matching above "three", and the teacher had to match the quantities for them to then glue to the paper. Another observation involved making an AB pattern with different paints, however the teacher did not provide any instruction nor point out the repeating unit "AB". Instead she had template of a red-blue-red-blue picture in the middle of each table, and children either copied it or used their own colors. Another observation involved children tracing numbers with paint (1-10) (no prior direction except "paint the numbers"), and then as children finished, the teacher asked each of them to verbally recite "1-10" the numbers on their paper, and then they would be allowed to go to center-time.

For the second category (one teacher), she was observed providing explicit instruction throughout the activity. She started out by asking children to problem solve why some shapes (e.g., triangle) could not fit into other shapes in the puzzle (e.g., square), and continued to provide instruction on identifying the sides of shapes. Children then used shape blocks and matching shape templates to create a picture.

Overall, I was pleased that children were receiving small-group math instruction. However there appeared to be a consistent lack of developmental appropriateness across all of the activities, despite the observed differences in teacher instructional support. First, a majority of children could not recognize numerals 1-10, nor successfully count one-to-one. Patterning instruction involved simple copying of AB patterns (without the teacher identifying the repeating unit), and most children finished the patterning activity in 2-3 minutes. And even in the one observation where the teacher offered open-ended questions and instruction throughout the activity, the instruction on sides of shapes seemed appropriate until she introduced counting the sides of a hexagon, when most children could not distinguish and name a triangle vs. square.

***Large group*.** Five out of 10 teachers included math during their observed large group time. However, all of these activities involved either identifying numbers on the calendar or counting attendance. In the teacher meeting, eight teachers (*N*=8) expressed concern about large-group math activities more generally and noted that their students would be more likely to focus in small-group and are not used to large-group instruction. For example, teachers reported typically utilizing a 10-minute large group for movement and song activities, morning meeting/calendar, and read-alouds.

Transitions. There were some instances of math during transitions and meal-times. For example, seven teachers were observed counting from 1-10 while children were washing their hands or brushing their teeth. And four teachers were observed counting out the number of cups or bowls for each child during meal-time.