GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Designing Studies to Test Causal Questions about Early Math: The Development of Making Pre-K Count

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Mattera, Shira	MDRC	
Morris, Pamela	New York University	
Jacob, Robin	University of Michigan	
Maier, Michelle	MDRC	
Rojas, Natalia	New York University	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) June 23, 2017

Check type of content being submitted and complete one of the following in the box below:

☐ If article: Name of journal, volume, and issue number if available
☐ If paper: Name of conference, date of conference, and place of conference
☑ If book chapter: Title of book, page range, publisher name and location
☐ If book: Publisher name and location
☐ If dissertation: Name of institution, type of degree, and department granting degree

DOI or URL to published work (if available) doi: 10.1016/bs.acdb.2017.04.002.

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

“This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scier through **[Grant number]** R305B1400 to **Institution]** New York Univ. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

Designing Studies to Test Causal Questions About Early Math: The Development of Making Pre-K Count

Shira K. Mattera[*,1], **Pamela A. Morris**[*,†], **Robin Jacob**[‡], **Michelle Maier**[*], **Natalia Rojas**[†]
[*]MDRC, New York, NY, United States
[†]New York University, New York, NY, United States
[‡]University of Michigan, Ann Arbor, MI, United States
[1]Corresponding author: e-mail address: shira.mattera@mdrc.org

Contents

Abstract

A growing literature has demonstrated that early math skills are associated with later outcomes for children. This research has generated interest in improving children's early math competencies as a pathway to improved outcomes for children in elementary school. The Making Pre-K Count study was designed to test the effects of an early math intervention for preschoolers. Its design was unique in that, in addition to causally testing the effects of early math skills, it also allowed for the examination of a number of additional questions about scale-up, the influence of contextual factors and the counterfactual environment, the mechanism of long-term fade-out, and the role of measurement in early childhood intervention findings. This chapter outlines some of the design

Advances in Child Development and Behavior
ISSN 0065-2407
http://dx.doi.org/10.1016/bs.acdb.2017.04.002

considerations and decisions put in place to create a rigorous test of the causal effects of early math skills that is also able to answer these questions in early childhood mathematics and intervention. The study serves as a potential model for how to advance science in the fields of preschool intervention and early mathematics.

Carefully conducted longitudinal research has led many researchers in the field to argue that preschool children's math competencies are a foundational outcome that may be a pathway to improving a broader set of outcomes for children into elementary school. More specifically, such research has suggested that math may be a vehicle to improving a broad set of children's competencies in addition to math, including language, early reading, and executive function (made up of several factors, including inhibition, cognitive flexibility, and working memory). And theory has been brought to bear to support these claims from longitudinal research: math may build language skills because math learning expands and enriches children's vocabulary—for example, when children learn about comparisons such as "more" and "less"—and math requires children to use language to express and to justify mathematical thinking (Ginsburg, Lee, & Boyd, 2008). In addition, the computational demands of math may build children's working memory and problem-solving skills, both components of children's executive function (Diamond, 2013; Duncan et al., 2007).

Although it is clear that early math skills are *associated* with later positive outcomes, less evidence exists about the long-term causal effects of intervening directly to improve children's math abilities. In other words, if you actively change children's early math skills, does that causally lead to changes in those associated long-term outcomes? A recent paper by Greg Duncan and colleagues, who conducted much of the meta-analytic work demonstrating the predictive power of early math, raises this exact issue (Bailey, Duncan, Odgers, & Yu, 2017). Analysis in the paper demonstrates that associations and quasi-experimental estimates of the relationship between early math and later outcomes are consistently larger than experimental estimates from smaller studies. As a result, Duncan and his colleagues (Bailey et al., 2017) urge the field to consider whether predictive associations between early skills and later outcomes should be interpreted as causal in nature. In doing so, they identify a number of explanations that would account for correlational associations between math and later outcomes. Although one plausible explanation for the correlation is that teaching and improving measureable early math skills directly affects later outcomes, others they propose would suggest that the correlation between early math skills and later outcomes is driven by one or more underlying internal or external stable traits, contextual factors, or aptitudes that account for both early and later outcomes (Bailey et al., 2017).

This chapter describes an effort, called Making Pre-K Count (MPC), that was launched to rigorously address the question about the long-term effects of early math skills. The large-scale randomized controlled trial of early math programs was unique in that it accomplished two goals. First, it contributed to the evidence base around the causal effects of the long-term, cross-domain associations that had been identified in the literature (Duncan et al., 2007). Causal tests that explore whether improving early math abilities at scale in the real world can improve later outcomes can help home in on math's potential as a target of intervention and a public policy lever. Second, in addition to causally testing the effects of early math skills, it also allowed for the examination of a number of additional questions about scaling up promising programs, the influence of contextual factors and the counterfactual environment, the mechanism of long-term fade-out, and the role of measurement in early childhood intervention findings. What differentiates this effort from others that preceded it is that it was intended *not* only as a test of the efficacy of a particular early childhood program. Rather, it was explicitly designed as a test of the predictive nature of early math skills across *domain*s and across *time*, at scale in a complex real-world *counterfactual* condition. In doing so, it may provide a model about the ways in which some of the most rigorous designs—randomized trials—might be brought to bear to test long-standing questions about the ability to scale-up programs in the real world, the role of context in program impacts, theories about developmental trajectories of learning and fade-out, and issues of measurement in our field—and, in this way, advance research more broadly in the field of early childhood education.

1. MAKING PRE-K COUNT AND HIGH 5s

The MPC and High 5s studies were led by the Robin Hood Foundation and MDRC, in collaboration with Bank Street College of Education and RTI International. MPC is the first study of the Robin Hood Early Childhood Research Initiative, which was designed to identify and to rigorously test promising early childhood interventions. The initiative is a partnership between Robin Hood, one of New York City's leading antipoverty organizations, and MDRC, a nonprofit, nonpartisan education, and social policy research organization. MPC is also supported with lead funding from the Heising-Simons Foundation, the Overdeck Family Foundation, and the Richard W. Goldman Family Foundation.

MPC tests whether implementing an evidence-based math curriculum (Building Blocks; Clements & Sarama, 2007) along with training and coaching of teachers would improve 4-year-olds' short- and long-term outcomes, relative to the typical pre-k math experience in New York City. This complex counterfactual condition provides the opportunity for a deeper examination of the role of contextual factors, such as classroom math practices and program and student characteristics, for program effects. High 5s builds on the MPC study, aligning children's math experiences from pre-k through the end of kindergarten in an attempt to sustain gains from the pre-k experience and learn about the role of fade-out in early childhood programming. High 5s provides small-group math "clubs" three times a week to some of the kindergartners who had received the MPC math curriculum the year before. The two studies rigorously examine the short- and long-term impacts of math interventions in preschool and kindergarten and also provided much-needed information about the role of measurement in the detection of effects across time. Follow-up data collection planned on a wide array of domains of child learning and development, through early elementary school, allows for the study of any short-term improvements of gains in math learning on children's long-term developmental trajectories.

2. DESIGN CONSIDERATIONS FOR BUILDING A STRONG TEST OF THE EFFECTS OF MATH ON CHILDREN'S LONG-TERM OUTCOMES

This chapter details the most consequential design elements that were put in place as part of the MPC study to build the strongest possible test of the strength of early math skills in the real world that is also able to answer new and emerging questions in early childhood mathematics and intervention. These include several features of the intervention and the counterfactual: (a) *identifying the most promising intervention* to be tested; (b) ensuring adequate *support for strong implementation* at scale; and (c) taking into account the "*business as usual*" *condition*. They also include a number of features of the design: (d) *the need for a rigorous random assignment design* that was unlikely to lead to differences in the intervention or control sites at the start of the study; (e) a large sample of children assessed in *multiple domains and followed well into elementary school*—enabling us to capture the longer term, cross-domain effects of the intervention and thereby gain some perspective on the likely impact of the intervention on later outcomes. Finally, the design included an innovative additional randomization study on top of the original one, to address the potential "fade-out" of effects. We address each of these in turn, later.

2.1 Selecting the Strongest Intervention

First and foremost, a key component of this study of early math skills required the selection of a program with strong potential to increase children's math learning. And, evidence of efficacy was needed across diverse sites and populations, especially populations that would most closely resemble those likely to be found in the selected site for this test—New York City. Secondarily, to ensure that the program could be well-implemented by a research team that were not program developers, the program that was selected needed to be ready for scale.

A structured review process was created for the purpose of selecting the math curriculum that could best meet the goals of the study. First, the team reviewed the early childhood education literature, read evidence-based practice guides, and spoke with experts in the early childhood education field to identify promising instructional models for preschool mathematics. Nine math instructional models were identified in 2011. To identify the strongest possible intervention to test, these potential math curricula for the study were evaluated in three key areas: (1) efficacy evidence, (2) curriculum content and program delivery, and (3) the implementation requirements and readiness for scale-up.

2.1.1 Review of Efficacy Evidence

The first step involved an assessment of the empirical evidence of efficacy, indicated by findings of at least one randomized controlled trial with preschool children; particular attention given to research with samples of children from low-income communities in New York City or other large urban areas, and to replication of findings across studies. The research literature on each of the nine math models was reviewed to determine whether there was sufficient evidence of efficacy to warrant further consideration for the first initiative. Based on this minimum standard of efficacy evidence, six of these models were excluded from consideration for the first initiative. Three remaining curricula had sufficient efficacy evidence for further consideration: *Building Blocks* (Clements & Sarama, 2007), *Big Math for Little Kids* (*BMLKs*; Ginsburg, Greenes, & Balfanz, 2003), and *Pre-K Mathematics*

(Klein, Starkey, & Ramirez, 2002). However, in a head-to-head test of *Building Blocks* and *Pre-K Mathematics*, *Building Blocks* had larger impacts. For that reason, only two early childhood math curricula were included in the more intensive review: *Building Blocks* (Clements & Sarama, 2007) and *BMLKs* (Ginsburg et al., 2003). See Table 1 for the review findings based on studies available in 2011.

The two remaining curricula were reviewed for more detailed evidence of efficacy for classrooms/teachers and for children. The review examined whether each curriculum's efficacy evidence from at least one large, multisite random assignment study took place in a major metropolitan area with a sample with characteristics similar to MPC's targeted sample (a high percentage of racial and ethnic-minority families and low-income families). Within the relevant studies, the review looked at whether there were significant effects on child math outcomes after preschool of at least 0.2 SD on a nationally normed measure or at least 0.5 SD on a developer measure. As can be seen in the top panel of Table 1, both programs had evidence of effectiveness for a relevant sample and context. Building Blocks had a larger evidence base, but BMLKs evidence came from a New York City-based trial.

2.1.2 Review of Content and Program Delivery

The second step included a review of curriculum materials for alignment with early math standards, strategic sequencing choices to maximize child learning, integrated assessment, and program delivery modes that were appropriate for the preschool models that were included in the study. The team conducted an extensive review of the lessons and activities in each curriculum, in order to understand the breadth of mathematical content as well as depth of coverage of each concept within the curriculum. Each curriculum was analyzed to identify the percentage of total curriculum time in which each math content area and each of its subtopics was addressed as a primary focus.

In addition, the curricula were reviewed to see whether the approach to program delivery and sequencing was based upon a clearly articulated theory of children's math learning development, the program emphasized teacher-guided learning with interactive participation of children, the program had an instructional software component that is integrated with the curriculum, and that assessments were integrated into the curriculum and used for leveling of activities and individualization. The second panel of Table 1 shows that, at the time, both programs met these requirements. *BMLKs* covered

Table 1 Review of Building Blocks and Big Math for Little Kids Pre-K Math Curricula

Criteria	**BMLKs**	**BB**
Efficacy evidence		
1. Efficacy evidence from at least one large, multisite random assignment study in a major metropolitan area, including a high percentage of minority families and low-income families	✓	✓+
2. At least one efficacy study found significant effects on child math outcomes after preschool of at least 0.2 SD on a nationally normed measure or at least 0.5 SD on a developer measure	a	✓+
Content and delivery		
3. Coverage of a range of math content areas in alignment with NCTM standards, including thorough instruction in basic early math skills (e.g., number, cardinality, shape), as well as complex early math skills (e.g., patterns, operations, measurement)	✓+	✓
4. The approach to program delivery and sequencing is based upon a clearly articulated theory of child development	✓	✓+
5. Program emphasizes teacher-guided learning with interactive participation of children	✓	✓
6. Program has an instructional software component that is integrated with the curriculum	b	✓
7. Assessments are integrated into the curriculum, and used for leveling of activities and individualization	✓	✓+
Implementation requirements and readiness for scale-up		
8. Program has an established professional development model	✓	✓
9. Meets all or most math-related standards in the New York State Early Learning Standards and Common Core Learning Standards for Prekindergarten, and also the Head Start Child Outcomes Framework and Head Start Performance Standards	✓	✓
10. Spanish language version of the curriculum is available	✓	c

Note: In this criteria table:

- A check mark (✓) indicates that the curriculum met the criterion in a satisfactory way
- A check plus mark (✓+) indicates that the curriculum exceeded the criterion, such as having efficacy evidence from multiple studies rather than just one
- A letter indicates that the curriculum did not meet the criterion fully, but there are footnotes below with additional information that is relevant to the criterion
  a. While *BMLKs* did not meet the second efficacy criterion for the preschool year, this criterion was met for the kindergarten version of the curriculum.
  b. *BMLKs* did not have integrated software in 2011, but it was in development.
  c. Spanish versions of the software and some curriculum materials were available, but the full *BB* curriculum had not been translated into Spanish.

a broader range of content areas than *BB*, specifically touching on each of the areas included in the NCTM math standards. *BB* focused more on number and operations and a bit less on some math topics such as measurement and representing data. Still, *BB* did cover most of the topics in the NCTM standards, and the developer had a developmental rationale for the emphasis on certain topic areas that were included in the curriculum. In addition, the sequencing of *BB* was based upon child learning trajectories in math content areas that extend through sixth grade, and topics are reinforced throughout the year as the level of difficulty progresses, whereas BMLKs relied on discrete units. At the time of the review, BMLKs also did not have a finalized computer program.

2.1.3 Review of Implementation Requirements and Readiness for Scale-Up

Finally, curriculum materials were reviewed and program developers were interviewed to determine whether the curriculum model was ready for large-scale implementation in the study context. A central question for this study was to examine the effects of a math intervention in the complex, real-world context of New York City pre-ks. Identifying curricula that not only focused on the appropriate math domains but also had the infrastructure to support teachers' practices in those math domains was essential for addressing questions about program implementation and scale-up. The review included information about the level of training, coaching, technical assistance, and other preparations necessary, and the presence of a standardized curriculum manual to achieve significant effects on teacher's math practices in a large-scale implementation, as well as the developer's availability to support such a large-scale implementation of this nature. The team also examined whether the curricula were appropriate for the study context (New York City) in meeting New York's math-related standards and having appropriate supports for the City's diverse student population by having Spanish-language materials available. Both programs had manualized curricular materials, although those from *BB* were ready for dissemination to teachers on a large scale. At the time of the review, Building Blocks had some Spanish-language materials but was not fully translated into Spanish. Also at the time, BMLKs did not have a manualized professional development model and had not previously used coaches to support implementation.

Based on this in-depth review, the team selected the *Building Blocks* math curriculum (Clements & Sarama, 2007) to include in its study of children's long-term math learning. The curriculum had strong efficacy evidence for teachers and children across a number of studies and diverse setting. In addition, the curriculum was sequenced in a way that experts believe may be important for child learning, cycling to reinforce content over the year while building difficulty over time. Finally, the program had a high level of manualization, with formal, written lesson-plans for teachers and documents to fully support the training, coaching, and technical assistance appropriate for use in diverse contexts. These three characteristics suggested that Building Blocks was highly likely to lead to changes in teachers' math practices and children's math outcomes in the preschool year when scaled-up across New York City preschools, allowing for a strong test of the effects of gains in math learning in preschool on math and other domains of learning in the short- and long-term and further investigation of variation in real-world implementation within the context of New York City.

2.2 Designing tor Scale: Ensuring Strong Implementation

The work reviewed in the curriculum selection stage demonstrated that curricular enhancements can be implemented effectively to produce moderate to large improvements in children's math skills, at least in the short term. Yet, despite this promising research, there is still a need for strong evidence that these gains can be achieved when the models are tested at scale amid the complexity of New York City's service delivery system. To support the Building Blocks curriculum in this study and address the question of scale-up, an implementation infrastructure was created in New York City. The infrastructure included ongoing training provided by the developer-approved trainers, in-classroom coaching provided by a local high-quality provider (Bank Street College of Education), and technical assistance supported by a management information system.

Developer-created training content was provided across 2 years to over 170 lead and assistant teachers. In the first year, lead and assistant teachers were offered 6 days of Beginner Training, which focused on understanding the curricular content and activities. In the second year, lead teachers were offered 5 days of Advanced Training, focusing on using child data to differentiate instruction to children at different skill levels. To address turnover between years, the Beginner Training was also reoffered in the second year to any new lead teachers. Curriculum developers had a small cadre of trained staff who had previously led teacher training. As part of scaling up, curriculum developers also trained additional local trainers to ensure there were enough skilled staff to provide training.

A coaching model was developed in collaboration with Bank Street, based on best practices in the field and the coaching model previously used to support the curriculum (TRIAD; Clements, Sarama, Spitler, Lange, & Wolfe, 2011). Pairing teacher training sessions with ongoing, in-classroom coaching has been shown to have the best effect on supporting teachers' transfer of what they have learned in training to their work with children in the classroom (Joyce & Showers, 2002; Sheridan, Edwards, Marvin, & Knoche, 2009). Bank Street hired, trained, and cosupervised the MPC coaches. The coaching model required that coaching occur in the classroom, when the coach could observe teachers implementing the Building Blocks curriculum and support implementation in real time as needed. Coaches also met with the lead and assistant teachers to reflect on how implementation had been going in the classroom that week and offer implementation support. In the first year, teachers received an average of 149 out of the expected 180 min of coaching weekly. In the second year, teachers received 99 out of the expected 120 min of coaching every other week. Nearly, all coaching sessions that were expected to occur took place (96% in the first year, 98% in the second year).

Finally, technical assistance was offered by MDRC, the research organization leading the study. A management information system was designed to track coaching and implementation dosage and quality through online logs completed by the coaches. Coach logs collected information about how often the Building Blocks activities were implemented in the classroom, the quality of each individual curricular component, and the overall quality of implementation. For example, the logs asked about how often Building Blocks whole-group activities were conducted that week and the quality of whole-group implementation as seen by the coach. The logs also asked more holistically about implementation, using Likert-scale type questions such as "This class looks like a Building Blocks classroom." The technical assistance team monitored these logs in real time, comparing data against a set of prespecified benchmarks created in collaboration with the curriculum developers. Table 2, drawn from the MPC pre-k report, outlines the prespecified benchmarks for each curricular component. For example, Building Blocks includes a whole-group math activity for each day of its 30-week curriculum. The whole-group benchmark specified that a core

Table 2 Building Blocks Curricular Component Definitions and Technical Assistance Benchmarks (Morris, Mattera, & Maier, 2016)

Curricular Component	Definition	Technical Assistance Weekly Benchmark
Main component		
Whole group	Activity led by a teacher and conducted with the majority of children in a class.	A core whole-group activity is completed most days (at least 66%) that children are in attendance in a week.
Small group[a]	Activity led by a teacher and conducted with three to four children in a class.	At least 75% of all children participate in small group.
Hands on math centers	Activities or manipulatives[b] for children to work and play with independently, or with a small group of children, with or without a teacher.	Math center activities are available most days (at least 66%) that children are in attendance in a week.
Computer	Activities made available to children through the BB web-based computer software.	At least 75% of all children participate in computer activities.
Supplementary component		
Small group record sheet	Template for teachers to record children's participation in and response to small group activities.	Teachers fully complete at least one small group record sheet for the week.
Family letters	Ready-made letters in English or Spanish that are sent home with children to help parents reinforce BB content at home.	Family letters "sent home" or "not sent home but did not need to" during the week.
ConnectED	Teachers' version of BB software that allows them to assign computer activities to children and to review reports of children's activity completion and progress.	Teachers access ConnectED during the week.

[a] Each week, the Building Blocks (BB) curriculum included one or two small group activities. In Year 1 (2013–2014), the research team asked teachers to conduct at least one small group activity on a weekly basis. In Year 2 (2014–2015), teachers were asked to conduct both small group activities if more than one was listed.

[b] Manipulatives are hands on objects that allow children to explore abstract math concepts concretely. Clements and Sarama (2013).

whole-group activity should be completed most days (at least 66%) that children are in attendance in a week. The benchmarks were designed to take into account developer expectations for implementation and help identify programs in need of additional support and technical assistance. The team met weekly with coaching staff to troubleshoot implementation challenges and to identify and provide support to persistently low-implementing classrooms.

These three components were essential to addressing the questions about early math *at scale*. The training and coaching put in place in this study was comparable to the implementation supports put in place in other studies of Building Blocks as well as real-world implementation of Building Blocks in Boston Public schools (Clements et al., 2011; Weiland & Yoshikawa, 2013). In other words, the implementation levels achieved by this implementation infrastructure created conditions similar to those that would be expected with real-world implementation outside of a research design. But most importantly for the causal effects of math—the model that was put in place was designed to have a good chance of changing math practices among preschool teachers as a vehicle to improving the early math skills of young children.

2.3 Considering the Counterfactual Condition and Context

A key consideration for developing a strong test of children's early math skills is ensuring that the intervention makes a difference over and above what children would normally receive in preschool—that is, over and above the counterfactual condition. The impact of an intervention is produced by the program's treatment contrast, which is the difference between the actual services received by participants and their counterfactual service receipt (what they would have received if they had not participated). That condition is defined by a combination of the program or services already in place, the characteristics of the participants, and the context in which these participants receive these services. In this case, given our interest in improving math skills, our focus was on ensuring that any program that was implemented in the program group classrooms would be better in terms of supporting math skills, than that which children would receive in the absence of the treatment.

In the past, math instruction has seemed particularly ripe for this kind of a test because math instruction has historically been underemphasized in preschool. Earlier descriptive studies have found that preschool teachers rated math as the lowest priority for children, behind preliteracy and social–emotional development (Lobman, Ryan, & McLaughlin, 2005). When researchers conducted focus groups with these same preschool teachers, teachers had "no substantive ideas about how teachers could be prepared to teach [math]" (Lobman et al., 2005). And, this is highly consistent with findings on classroom instruction: research has shown that when teachers do provide any lessons in math, they have often focused on quite simple aspects, such as the names of shapes and numbers from 1 to 20, without the richness of mathematical reasoning, inferences, metacognition (explaining how one arrived at an answer), and complex vocabulary that characterize many of the successful, developmentally appropriate curricula that experts recommend (Ginsburg et al., 2008; National Research Council, 2009). Clifford et al. (2005) found that only 8% of classroom time was spent on math activities involving counting, time, shapes, sorting, while 21% was devoted to literacy activities. Thus, it was thought that providing teachers with training in delivering more math instruction might represent a substantial shift in children's preschool experiences compared with what the typical preschool has provided (and thus a resulting change in children's math learning).

Interestingly, "business as usual" pre-k programming in the MPC study turned out to be a shifting target. The years of the study (2012–2015), including a pilot and 2 years of implementation, took place during a time of major change for New York City pre-K. In 2011, the city began rolling out their preschool Common Core learning standards, requiring programs to implement a Core-aligned curriculum of their choice and implementing math and literacy tasks embedded in thematic units. In 2012, the city established consistent program quality requirements across all sites through EarlyLearn, its consolidation of funding streams. And in 2014, Mayor Bill deBlasio's signature campaign promise of full-day pre-k for all 4-year-old children led to the creation of an additional 34,000 pre-k slots.

Although none of these initiatives was a direct math intervention, they led to growing scrutiny of pre-k and a focus on instructional quality. This context was different than the pre-k landscape of 10 years ago. While in theory this may or may not have led to growing math content being delivered in preschool classrooms, in practice, this translated into a *much larger* amount of math—35 min of math instruction—taking place in New York City pre-ks compared to prior studies of Building Blocks (and, indeed, other math-focused preschool curricula, as well). This context, which included a much greater amount of time on math than has been seen in other studies, suggests that the nature of math instruction may be changing and math interventions may have a higher bar to clear to make a difference for young children.

A key aspect of understanding this context was data collected in the study on teachers' math practices, classroom climate, and the structure and content of instructional time during the day. These data were used to measure impacts on teacher practices and classroom processes, and therefore were collected not only in the program group, but in the control group as well. Collecting detailed, longitudinal data about the counterfactual condition was a critical component of the study—because it provided not just an analysis of program impacts but a deeper examination of the ways in which the historical and geographical context for the study may have played a role in producing such program impacts.

2.4 Building a Rigorous Design

To investigate the relationship between early math and later math, language, and executive function outcomes, it was critical to implement a rigorous causal design. As such, a cluster-randomized controlled trial was implemented, in which full pre-k centers were either offered the math program or assigned to a control group using a lottery-like process. This randomization process (a) ensured minimal, if any, spill-over from one group of teachers to another (which was especially important for this intervention, which calls for teachers to implement the program for 2 years in order to get comfortable with the program); (b) strengthened the impacts on teacher practice by espousing "whole school reform" efforts; and (c) most importantly, allowed for the calculation of unbiased estimates of treatment that can be leveraged for other causal estimates of the effects of math on other aspects of child learning and development (Bloom, Bos, & Lee, 1999; Gennetian, Magnuson, & Morris, 2008).

Thirty-five of 69 pre-k sites serving predominantly low-income children were assigned to receive the math curriculum, training, and coaching (the program group) over 2 years. Teachers in these 35 program sites spent the first "gentle introduction" year learning the curricular components and activities and receiving weekly coaching to support classroom implementation, with the second implementation year focused on high-quality implementation. The other 34 sites were assigned to continue their typical programming (the "pre-k-as-usual" control group) during the same 2 years. While teachers implemented the curriculum and received coaching and training for 2 years, as designed by the program developer, impacts on child outcomes were assessed for the children who were in the classrooms in the second year of implementation. This design allowed teachers to become familiar with the program, while still testing the effects of 1 year of the Building Blocks curriculum for children.

Notably, we blocked pre-k sites into groups of four to six sites before randomization. This blocking was critical to ensure a strong "match" between treatment and control sites in a constrained sample. We did not want the lottery-like process of randomization to result in all program sites in a single borough or serving a particular demographic group of children, and all program sites in another or serving a different demographic group, as we discuss later. Sites were selected to reflect the geographical, racial, and ethnic diversity of New York City's low-income population. Approximately two-thirds ($n = 47$) of sites were in public schools, while the remainder ($n = 22$) were in community-based organizations. Sites were selected in four of New York City's five boroughs. Sites were blocked based on three criteria: venue (CBO vs public school), borough, and racial-ethnic composition (sites that served a primarily Hispanic population vs other). Blocking of sites was conducted for two key reasons: first, it provided for some protection against a "bad" draw in which treatment and control sites are poorly matched. This is particularly likely to occur in studies such as this one in which there are few units at the level of randomization (i.e., only 60–70 sites). Second, blocking (as compared to "pairing" sites) provided some protection against the loss of sites between randomization and the beginning of the "impact" or second year of study. That is, given the lengthy period of implementation, sites could have dropped out of the study before the second year of implementation—although in practice, this did not happen. Given the challenge of holding onto a matched waitlist of sites for a lengthy period, blocks of four or larger sites made it such that the block could be part of the analysis even if a single center had dropped out of the study.

The blocked, cluster-randomized controlled trial spanning 2 years of implementation ensured the most rigorous test of the effects of an early math intervention. Blocking sites to account for a 2-year implementation maximized teachers' time and familiarity with the curriculum by allowing them time to grow comfortable with implementing the curriculum (and thus ensuring strong potential impacts on teachers' practice—the key driver to strong impacts on children's math learning), while protecting the random assignment design against the potential for attrition across such a long study window. Randomly assigning full pre-k sites minimized the likelihood of spillover from one classroom to another while also capitalizing on training and coaching cohorts of teachers within a school.

2.5 Measuring Children's Outcomes Over Time

One reason improving math has seemed promising as an intervention strategy is the set of findings from longitudinal studies suggesting a link between early math skills and children's later outcomes (Duncan et al., 2007; Duncan & Magnuson, 2011). Seminal studies such as Perry Preschool or Abecedarian demonstrate long-term impacts of early cognitive intervention, even in light of early effects "fading out." Some researchers have suggested that the relationship between impacts on early cognitive skills and later reemerging impacts in adulthood may pass through noncognitive or "dark matter" skills that have historically not been measured, such as executive function (Heckman, 2006). However, this theory has not been adequately tested for two primary reasons. First, few intervention studies follow children over time, allowing for an examination of program impacts longitudinally. Additionally, measurement challenges in the field of early childhood make it difficult to parse apart whether a comprehensive set of developmental domains have been assessed at all vs whether they have been assessed well enough to measure program effects.

To test the developmental theory that improving math skills could spill over into children's other outcomes and to examine the role of measurement in the detection of those effects over time, it was imperative to design the study (a) to examine children's outcomes in the short- and long-term and (b) to include a broad, rich set of measures that would allow for a test of math on children's learning in other domains of development and an examination of the role of measurement on previous patterns of program impacts. Such a longitudinal data collection would collect rich data at multiple follow-up time points to fully understand the long-term effects of supporting math skills. The data collected would also include multiple measures within the same domain to parse apart how measurement influences the ability to detect program effects. The MPC study addressed this issue, with a data collection plan designed to collect direct child assessments and school records across a variety of outcome domains from preschool through third grade.

Described in greater detail in the next section, research on early childhood programs shows that effects on some outcomes may fade over time, but long-term impacts on income can sometimes still be observed into adulthood (Chetty et al., 2011). This suggests that certain unmeasured variables may in fact be a pathway through which interventions affect long-term outcomes. An important theoretical reason for this study's focus on math was not only to bolster children's math skills in the short term, but also to assess whether improvements in math capabilities would translate into improvements in other domains of learning and development and across time. Indeed, when examining the effects of an intervention, many studies only assess the targeted domain of interest—in this case, math—limiting our ability to address questions about the cross-domain implications of early math skills. In MPC, we intentionally addressed this gap, by collecting information both on outcomes directly related to the intervention (such as math skills) and other key domains (language, executive function) that are of critical importance to advance the field's understanding of preschool's effects into elementary school, and eventually adolescence and adulthood.

In identifying the constructs to be assessed, two key issues were considered: (1) which constructs were most aligned with the intervention's focus, directly as well as indirectly, and (2) which constructs matter most for children's later outcomes. A careful review of the program's theory of change was used to prioritize constructs aligned with the intervention's focus, while a review of longitudinal research helped identify constructs that matter most for children's outcomes at different time points. For example, while receptive language was an important indicator of a child's understanding of rich language and math talk in the early years, by third grade the project's language focus had shifted to reading comprehension, given the math intervention's focus on children's critical thinking skills.

Measuring children's skills across time presents a number of design challenges. First, children's skills in the preschool-to-third-grade age range change exceptionally fast. As with every longitudinal study, it may not be possible to collect the same measures across multiple time points. For instance, executive function measures of inhibition in preschool (such as a computerized pointing task or a Simon Says-like task) are too easy for third graders and must be adapted for later skills. To address this issue, the MPC study first identified the most promising measures for each construct of interest in preschool. Then, the measures were mapped across time to identify instruments either that could be used across multiple age ranges or that could be mathematically equated across time. For measures or domains where it was not possible to maintain the same instrument across time, the study identified new measures that used similar methodologies and captured the same construct if possible.

In pre-k, the study assessed children's math skills with two direct assessments. The Early Childhood Longitudinal Study-Birth Cohort (ECLS-B; Najarian, Snow, Lennon, & Kinsey, 2010) is a more proximal, IRT-based math measure assessing children's number sense, operations, measurement, geometry, spatial sense, and patterns. The Woodcock-Johnson III Tests of Achievement: Applied Problems (WJ-AP; Woodcock, McGrew, & Mather, 2001) is a more distal, nationally normed math measure of mathematical thinking. Importantly, different measures of the same construct may provide unique information. For example, the study was interested in two aspects in the math domain. A detailed and sensitive measure of children's math ability in multiple subdomains (e.g., numeracy, geometry, patterning)—the ECLS-B—was needed to clearly identify the areas of math where the program had effects. At the same time, a more widely used but less sensitive measure with extensive literature on its predictive power—the Woodcock-Johnson Applied Problems—was needed to better understand what these impacts might mean for future outcomes. Including two differing measures of the same construct, each with differing strengths for assessing program impact and for assessing the long-term implications of such program impacts, strengthened the study substantially.

As indicated earlier, it was also critical to assess other domains of child learning and development in MPC. Receptive language was assessed using the nationally normed Receptive One-Word Picture Vocabulary Test (Martin & Brownell, 2011). Children's executive function (including working memory, inhibition, and cognitive flexibility) was assessed using three measures, each that tapped differing subcomponent skills of executive function. The Pencil Tap (Diamond & Taylor, 1996; Luria, 1966) assesses inhibition, the Spatial Conflict Arrows task (Willoughby, Wirth, Blair, & Family Life Project Investigators, 2012) assesses cognitive flexibility, and Corsi Blocks (Corsi, 1972; Lezak, 1983) assesses working memory.

Collecting information about program impact with differing measures of the same construct, across developmental domains, and across time allows for the most complete assessment of a program's impacts. This may be particularly important for an early childhood math intervention, which is theorized to have not only localized effects on math during the intervention year but also effects that spill over into other domains and across time. The measurement challenges encountered provided an opportunity for new learning into measure sensitivity and overlap, measurement equivalence, and how measurement influences the detection of program impacts.

2.6 Addressing the Fade-Out of Effects

One of the challenges in preschool intervention research is the potential for the "fade-out" of effects (Lee & Loeb, 1995). Frequently, the cognitive and academic benefits from even the most promising preschool interventions have not lasted when the children were assessed later in their elementary school years. As a case in point, MDRC's Foundations of Learning Project found benefits to preschool classrooms and children's outcomes (reduced conflict and increased classroom engagement) from a program targeting teachers' classroom management skills (Morris, Raver, Millenky, Jones, & Lloyd, 2010). However, the effects observed in preschool classrooms were not sustained after children entered elementary school classrooms. Such a phenomenon could be due to either the quality of the elementary schools that children ultimately attend, or to control group "catch up," but it suggests that ongoing support beyond a 1-year preschool intervention may be critical for sustained impacts. A follow-up to one of the Building Blocks trials added on a light-touch follow-through in kindergarten for some children who had received the Building Blocks curriculum in preschool. The follow-through taught teachers about the entering children's skills and some activities that could help support those skills. The follow-up showed that children who received the kindergarten follow-through in addition to preschool Building Blocks had less fade-out of effects in first grade than children who received Building Blocks alone (Sarama, Lange, Clements, & Wolfe, 2012).

In light of this, we wanted to take a multipronged approach to skill development that would include not only a strong approach to changing "pivotal" skills in preschool (i.e., our focus on math for MPC) but also an additional investment in improving children's experiences during kindergarten through ongoing intervention. From our perspective, to do the former without the latter would be insufficient. Moreover, this dual-pronged approach expanded the ability of MPC to answer additional questions about the course of child impacts and the potential fade-out of effects.

In designing this additional investment in kindergarten, we reviewed the literature and had a number of discussions with experts in the field. In doing so, we considered a number of different intervention options for implementation in kindergarten as an add-on to the MPC effort, including changes to the kindergarten curriculum, teacher professional development, in class enrichment, one-on-one tutoring, and pull-out small group enrichment. These options were evaluated against the following six criteria:

1. *Potential to sustain impacts*—the extent to which the program, when offered in kindergarten, had the potential to lead to continued math gains for the students who received the Building Blocks curriculum in preschool.
2. *Evidence of effectiveness*—the degree to which there was evidence to support the effectiveness of the program model.
3. *Logistical complexity*—the ease or difficulty of achieving buy-in from administrators and teachers, scheduling and logistics, and managing tutors/facilitators/coaches/teachers.
4. *Direct program intervention cost*—the degree to which the program had the potential to be cost effective. We considered the costs of developing or modifying an intervention, buying, or developing training materials, employing tutors/coaches/teachers, and providing them with training and supervision. We also considered possible additional implementation costs such as the extent to which technical assistance would be needed.
5. *Long-term scalability*—the likelihood that the program could be scaled in New York City or replicated in other areas.
6. *Maintain integrity of the MPC study design*—the degree to which the design for evaluating an add-on kindergarten intervention would maintain the original design for evaluating the pre-kindergarten math enhancement.

Based on these criteria, pull-out small group enrichment, or what we called a "math club" model, emerged as the most promising approach to support children's ongoing math learning following pre-k. A large number of studies of tutoring for low-achieving students have shown it to be an effective strategy for improving children's skills across domains, although most tutoring programs to date have focused on the literacy domain. In fact, a review found that, across 16 studies, the effect sizes for one-on-one tutoring were positive *in nearly every case* (Wasik & Slavin, 1993). This consistency of findings is quite unusual for intervention research. A meta-analysis conducted by Slavin, Lake, Davis, and Madden (2011) found an overall weighted mean effect size of 0.39 for one-on-one tutoring interventions by teachers, and a similar effect size of 0.31 for small group tutoring. Studies by Brown, Morris, and Fields (2005) and Ehri, Dreyer, Flugman, and Gross (2007) compared certified and paraprofessional tutors and found certified tutors to be much more effective, but both studies still found substantial positive effects for the paraprofessionals. And meta-analyses of tutoring with paraprofessionals shows positive results with struggling readers (weighted mean ES = 0.38; Slavin et al., 2011). The support for the math club option rested heavily on both its feasibility, its lower cost compared to a one-on-one tutoring intervention, and on the consensus that it would maximize the likelihood of large and sustained impacts on children's math skills because math clubs would provide an intensive, direct intervention for the children who had benefited from the *Building Blocks* curriculum.

Previous research demonstrates that high-quality math curricula are effective at raising the achievement of young children (Agodini, Harris, Thomas, Murphy, & Gallagher, 2010; Agodini et al., 2009; Hofer, Lipsey, Dong, & Farran, 2013; Klein, Starkey, Clements, Sarama, & Iyer, 2008; Sarama, Clements, Starkey, Klein, & Wakeley, 2008), and we designed the math club model to incorporate key elements that have been found to be associated with effectiveness including highly engaging activities and differentiated instruction in small groups.

The math club program was developed at the University of Michigan. Called *High 5s*, the program provided a small group enrichment experience for kindergarten students who had received the *Building Blocks* program in kindergarten, using staff who were offered a paraprofessional salary to implement the program. In the *High 5s* program, groups of approximately four kindergarten students met three times a week with a trained facilitator from Bank Street College of Education. Facilitators provided students with targeted instruction as they played fun, engaging math games in a small group setting. The High 5s program was developed with input from the developers of *Building Blocks* and was developed specifically as a follow on to the *Building Blocks* program, building on the "learning trajectories" which form the basis of Building Blocks. It was also designed to be aligned with the Common Core curriculum used by the New York City public schools, with the idea that the activities in the *High 5s* clubs would reinforce concepts learned in the classroom, but also go beyond the content typically covered in kindergarten. Activities were designed to move students along four key mathematical learning trajectories: counting, composition of numbers, early addition and subtraction, and geometry, and also included activities related to measurement and patterning. The program was intended to provide enrichment, not remediation, and to provide students with continued exposure to high-quality math instruction during the kindergarten year.

Once we had identified an approach to this additional boost in math in kindergarten, the next question was whether this should be provided to all children who had experienced enhanced math in pre-k or only a subset of children. In the end, a decision was made to rerandomize the children who had received the preschool math program—providing an embedded evaluation within the larger evaluation—to provide the greatest learning from the overall effort.

More specifically, the design of the kindergarten study was as follows: Students attending the 24 public schools participating in the MPC project who received *Building Blocks* during pre-k were eligible to participate in the High 5s programs. Once parental permission was obtained, eligible students were randomly assigned within schools to either receive the High 5s program or to a "business as usual" control group. A total of 25 facilitators, supported by five supervisors, administered the program to over 300 children in 79 clubs across 24 New York City public schools during the 2015–2016 school year.

This two-stage random assignment design allowed for a wider range of developmental questions to be addressed. The second round of random assignment creates three sets of causal comparisons: (a) the effect of receiving MPC relative to the typical pre-k experience in New York City; (b) the effect of receiving High 5s and MPC compared to MPC alone; and (c) the effect of receiving High 5s and MPC compared to the typical pre-k and kindergarten experiences in New York City. In addition to testing the direct effect of a new, small-group math program (question b), it also addresses questions about the developmental progression of children's skills and how those skills develop in different contexts. The study can examine how sustaining an aligned, enriched math environment over 2 years may affect the course of children's outcomes compared to 1 year of enrichment (or none). Noncausally, High 5s may also let us examine how or if impacts fade or outcomes grow across higher quality and lower quality instructional contexts. Directly speaking to the ongoing debate about the implications of program fade-out in early childhood, the MPC/High 5s two-stage random assignment design will provide much-needed descriptive, correlational, and causal information about the developmental course and sequelae of potential program fade-out.

3. EARLY FINDINGS AND CONCLUSION

Previous research makes it clear that math in the early childhood years could be a potentially meaningful lever for changing children's outcomes in the long term. However, whether a math intervention, implemented *at scale* in a complex and changing *control context*, could effectively leverage these associations between early math and later outcomes to have an impact, *measured across domains* and *sustained across time* for children's outcomes remained to be seen. The MPC and High 5s studies were designed specifically to test these questions. As such, the studies included a number of components to ensure a thorough and comprehensive examination of not just the causal effects of math but also implementation at scale, the role of classroom math context, the developmental progression of fade-out, and the influence of measurement.

This design has already proved fruitful in adding information to the field: Results in the pre-k year demonstrate that while teachers were able to substantially change the amount of math provided to preschoolers—and to a lesser degree, the quality of that math instruction—there were no effects on children's math, language, or executive function outcomes by the end of pre-k (Morris et al., 2016)—although there was evidence of early math impacts in the first half of the pre-k year (that is, based on a fall assessment, that does not appear to be due to a "poorly matched" random assignment process). This in itself is surprising, suggesting that more exploration is needed of the causal relationship between instructional dosage and quality and child outcomes and the role that the counterfactual condition may play in classroom-level and child-level impacts. Future research on early childhood intervention will need to begin to unpack causally which aspects of instructional quality matter, in which contexts, and how, for children's outcomes.

The larger challenge, of course, is that the lack of impacts on math learning at the end of pre-k may limit the potential of this study to address the larger questions about how early math learning might improve outcomes for children, across domains of learning. That said, the jury is still out on these findings. We have yet to fully investigate whether there are some groups of sites, teachers, or children for whom the implementation of Building Blocks produced positive outcomes for children, as well as whether the lack of impacts at the end of pre-k is due to children's "lack" of learning, to the insensitivity of our measurement tools to address children's deeper math learning, or to earlier-than-expected impacts in the *fall* of the pre-k year fading out faster than has previously been documented.

The design considerations laid out in this chapter make this study uniquely suited to shed light on all of these questions. Comprehensive information about the classroom context and student and teacher characteristics will allow for a nuanced exploration of the program's effects for different groups and contexts. The rich data collection reaching into kindergarten and elementary school will provide additional information about how different measures capture children's knowledge and learning and whether more sensitive measurement might detect program effects in a more nuanced manner. Additionally, the longitudinal nature of the data collection will add much-needed information about the trajectories of children's outcomes—and impacts—over time. This will allow for an investigation of whether fall impacts were an anomaly or an example of extremely early impacts and faster-than-expected fade-out by the end of pre-k.

Interestingly, new preliminary results examining the cumulative impacts of MPC and High 5s on children's kindergarten outcomes suggest that there *is* a positive effect of early math intervention on children's math and attitudes toward math in Kindergarten, 1 year later (although it is not yet clear which program contributed to which outcomes and how much; Mattera & Morris, 2017). Moreover, not only do positive impacts on math learning emerge, but executive function skills are also improved as a result of the enhanced math experiences across preschool and kindergarten. Although further analysis is needed to determine whether the effects are due to math learning in preschool vs preschool and kindergarten, the impacts act as a proof of concept to demonstrate that intervening with math in the early years can improve children's math and executive function outcomes, as well as affect how children feel about their math skills.

The unexpected results in pre-k and early kindergarten findings suggest that there is a lot we still do not know about how changes in math instruction affect children's outcomes in the long-term. MPC and High 5s demonstrate that rigorous causal tests, necessary to understand the effects of potential policy levers, can also provide rich descriptive information and a mechanism for investigating additional questions of interest to the field.

Even with considerably more investigation to be done, this and related work is critically important if our interest is in improving the field's knowledge to address the so-called achievement gap—the lower levels of achievement that follow low-income children relative to their more affluent peers as they move through school. If the causal relationship between math and later outcomes exists, then math may be a potentially powerful lever for change if our goal is to dramatically reduce the elementary school achievement gap. At the same time, there is considerable debate about the unique importance of which domain of school readiness matters and is most amenable to curricular intervention at scale, for whom, and in what context (Claessens, Duncan, & Engel, 2009; Duncan et al., 2007; Grimm, Steele, Mashburn, Burchinal, & Pianta, 2010). At the same time, these questions of the focus on effectiveness of early childhood intervention overlap with issues of the longitudinal mechanism of fade-out and measurement specificity and sensitivity. As a field, it is imperative that we begin to build a body of research to address these questions. The design considerations outlined earlier to create a rigorous and richly designed study provide one model to build such a body of research.

REFERENCES

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools. NCEE 2009-4052*. National Center for Education Evaluation and Regional Assistance.

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders*. Princeton, NJ: Mathematica Policy Research.

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39.

Bloom, H. S., Bos, J. M., & Lee, S. W. (1999). Using cluster random assignment to measure program impacts: Statistical implications for the evaluation of education programs. *Evaluation Review*, *23*(4), 445–469.

Brown, K. J., Morris, D., & Fields, M. (2005). Intervention after grade 1: Serving increased numbers of struggling readers effectively. *Journal of Literacy Research*, *37*(1), 61–94.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415–427.

Clements, D. H., & Sarama, J. (2007). *Building blocks: SRA real math, grade PreK*. Columbus, OH: SRA-McGraw Hill.

Clements, D. H., & Sarama, J. (2013). *Building blocks (Vols. 1 and 2)*. Columbus: McGraw-Hill Education.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*(2), 127.

Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., & Pianta, R. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*(3), 126–143.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Unpublished doctoral dissertation Montreal, Canada: McGill University.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "do as I say, not as I do" *Developmental Psychobiology*, *29*(4), 315–334.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428.

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In R. J. Murnane & G. J. Duncan (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 47–70). New York: Russell Sage Foundation.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*(2), 414–448.

Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, *44*(2), 381.

Ginsburg, H. P., Greenes, C., & Balfanz, R. (2003). *Big math for little kids*. Parsippany, NJ: Dale Seymour Publications.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. In *Social policy report*: *Vol. 22*. Ann Arbor, MI: Society for Research in Child Development. Number 1.

Grimm, K. J., Steele, J. S., Mashburn, A. J., Burchinal, M., & Pianta, R. C. (2010). Early behavioral associations of achievement trajectories. *Developmental Psychology*, *46*(5), 976.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902.

Hofer, K. G., Lipsey, M. W., Dong, N., & Farran, D. (2013). *Results of the early math project: Scale-up cross-site results*. Nashville, TN: Vanderbilt University. Retrieved from my.vanderbilt.edu/mathfollowup/reports/technicalreports.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

Klein, A., Starkey, P., & Ramirez, A. (2002). *Pre-K mathematics curriculum: Early childhood*. Glenview, IL: Scott Foresman.

Lee, V. E., & Loeb, S. (1995). Where do head start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis*, *17*(1), 62–82.

Lezak, M. D. (1983). *Neuropsychological assessments*. New York: Oxford University Press.

Lobman, C., Ryan, S., & McLaughlin, J. (2005). In *Toward a unified system of early childhood teacher education and professional development: Conversations with stakeholders Annual meeting of the American educational research association, Montreal, April* (pp. 11–15).

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Martin, N., & Brownell, R. (2011). *Receptive one-word picture vocabulary test*. Torrence, CA: Western Psychological Services.

Mattera, S. K., & Morris, P. A. (2017). *Counting on early math skills: Preliminary kindergarten impacts of the making pre-K Count and high 5s programs*. New York: MDRC.

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

Morris, P., Raver, C. C., Millenky, M., Jones, S., & Lloyd, C. M. (2010). *Making preschool more productive: How classroom management training can help teachers*. New York, NY: MDRC.

Najarian, M., Snow, K., Lennon, J., & Kinsey, S. (2010). *Early childhood longitudinal study, birth cohort (ECLS), preschool-kindergarten 2007 psychometric report (NCES 2010–009)*. Washington, DC: National Center for Education Statistics, Institute for Education Sciences, US Dept of Education.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academy Press.

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *1*(2), 89–119.

Sarama, J., Lange, A. A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly*, *27*(3), 489–502.

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.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1–26.

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 179–200.

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.

Willoughby, M. T., Wirth, R. J., Blair, C. B., & Family Life Project Investigators. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment*, *24*(2), 418.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities* (pp. 371–401). Itasca, IL: Riverside Pub.