CHILD DEVELOPMENT

Child Development, July/August 2021, Volume 92, Number 4, Pages e599–e620

Is Skill Type the Key to the PreK Fadeout Puzzle? Differential Associations Between Enrollment in PreK and Constrained and Unconstrained Skills Across Kindergarten

Meghan McCormick
MDRC

Christina Weiland
University of Michigan

JoAnn Hsueh, and Mirjana Pralica
MDRC

Amanda K. Weissman, and Lillie Moffett
University of Michigan

Catherine Snow
Harvard Graduate School of Education

Jason Sachs
Boston Public Schools

This study examines whether associations between enrollment in public and non-public PreK and children's (N = 508; M_{age} = 5.60 years in fall of kindergarten) math and language and literacy outcomes were more likely to be sustained through the spring of kindergarten for unconstrained versus constrained skills. Associations between public PreK and language, literacy, and math outcomes were more strongly sustained through the spring of kindergarten for unconstrained skills, relative to constrained skills. Only associations between non-public PreK and unconstrained language skills were sustained through the spring of kindergarten. Associations in the fall of kindergarten differed by family income and dual language learner (DLL) status but there was no subgroup variation by the spring of kindergarten. Implications for policy and practice are discussed.

Children who enroll in PreK programs typically demonstrate better cognitive ability, attention, language, executive functioning, and social-emotional skills at the start of kindergarten than their non-PreK attending peers (Phillips, Johnson, Weiland, & Hutchison, 2017). Yet, studies that have followed PreK attenders through elementary school generally show that impacts on cognitive and academic skills tend to diminish in early elementary school—a phenomenon commonly known as *fadeout* (wherein impacts of PreK programs fadeout over time) or *convergence* (wherein the children who are not assigned to the PreK program catch up to their PreK-attending peers over time; McCormick, Hsueh, Weiland, & Bangser, 2017). For example, findings from rigorous evaluations including the Head Start Impact Study and the Tennessee Voluntary Prekindergarten Effectiveness Study (Lipsey, Farran, & Durkin, 2018; Puma et al., 2010) revealed that PreK participants generally demonstrated higher levels of language, literacy, and math skills than their non-PreK attending peers prior to or at the start of kindergarten. However, these differences became much smaller and then ceased to exist as children moved through early elementary school. This trend is apparent across studies of PreK programs, with a recent meta-analysis finding

The research reported here was conducted as a part of a study funded by Arnold Ventures and Grant R305N160018—17 from the Institute of Education Sciences to MDRC with subcontracts to the University of Michigan, the Boston Public Schools (BPS), and the Harvard Graduate School of Education. Work done by Amanda Ketner Weissman and Lillie Moffett was supported by the Institute of Education Sciences Predoctoral Fellowship at the University of Michigan (Grant R305B150012). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Thanks to the BPS, Brian Gold, Blaire Horner, Annie Taylor, the BPS Department of Early Childhood coaches and staff, the BPS Department of Research, the MDRC team (Jennifer Yeaton, Kelly Terlizzi, Desiree Alderson, Rama Hagos, Marissa Strassberger and Sharon Huang), the Harvard Graduate School of Education research team (Sibyl Holland, Maia Gokhale, and field-based data collection staff), and the University of Michigan research team (Paola Guerrero Rosada, Kehui Zhang, Christina Daniel, and Margaret Michalowski). We also appreciate the thoughtful comments we received on this paper when we presented it for the IES Early Learning Network, at the first meeting of the Consortium on Sustaining PreK Effects, the Education Studies Colloquium at the University of Michigan, and the Stanford CEPR seminar. The Institutional Review Boards at the lead and partner organizations for this study approved the human subjects plan prior to the commencement of study activities. The project name is ExCEL P-3 and the IRB approval number is 860661-15.

Correspondence concerning this article should be addressed to Meghan McCormick, MDRC, 200 Vesey St. 23rd Floor, New York, NY 10281. Electronic mail may be sent to meghan.mccormick@mdrc.org.

© 2021 Society for Research in Child Development
All rights reserved. 0009-3920/2021/9204-0042
DOI: 10.1111/cdev.13520

that about half of the eventual convergence on cognitive outcomes occurs during kindergarten and then by about half again by the end of second grade (Li et al., 2016).

Constrained and Unconstrained Skills

Although there are a number of hypotheses explaining this phenomenon (e.g., Abenavoli, 2019; McCormick et al., 2017), one understudied theory motivated by work on constrained skill theory done by Paris (2005) is that PreK programs in general may focus more time on teaching and assessing children's constrained skills compared to unconstrained skills and effects of PreK on constrained skills may be less likely to be sustained across time. Constrained skills support students' development of instrumental competencies that are directly teachable and can be readily assessed. These are also skills that most children must master during elementary school in order to continue moving successfully through schooling (Paris, 2005; Snow & Matthews, 2016). The concept of constrained skills was originally developed in the domain of literacy and language. Teaching the alphabet, writing letters, practicing letter sounds, and spelling one's name would be considered constrained literacy activities. In this study, we also extend and apply concepts of constrained skill theory to math knowledge and skills. We argue that there are similarities in the definitions of rote and procedural math knowledge and constrained skills (Rittle-Johnson & Schneider, 2015). Number recognition, counting, arithmetic (e.g., addition, subtraction, multiplication, and division), and identification of shapes are examples of constrained math activities.

In contrast, unconstrained skills refer to a broader set of competencies that develop over time, will continue to develop over the full life course and can never be fully mastered, and are more complex and difficult to assess (Paris, 2005). Unconstrained language skills would include vocabulary and reading comprehension (Snow & Matthews, 2016), while unconstrained math skills would include algebraic thinking, composing shapes, patterning, comparing numbers, relational thinking, and applying strategic thinking to math problems. Building on work by Rittle-Johnson and Schneider (2015), we argue that higher-order conceptual representation of math knowledge is aligned with unconstrained skills. The focus here is not necessarily on being able to achieve a correct answer on an assessment, but rather, being able to use a variety of problem-solving and critical thinking approaches in coming to an answer.

Constrained and unconstrained skills are not necessarily separate and distinct categories and are not in conflict with one another. For example, reading comprehension can be thought of as relatively constrained if factual recall is the goal ("How many apples did the very hungry caterpillar eat?" or "How did Jay Gatsby meet Daisy?") or more unconstrained if it requires interpretation and speculation ("What are some good ways to get over a tummy ache?", "What does Daisy represent to Gatsby?"). With respect to math, third grade students might be taught or expected to be able to recite the times table up to 12 ($12 \times 12 = 144$). While *memorizing* multiplication tables is certainly a useful procedural skill that also supports the development of conceptual skills, it would be considered constrained in this framework because it gives rise to no capacity or insight beyond getting the right answer. Building on work examining the development of procedural and conceptual math skills (e.g., Rittle-Johnson, Schneider, & Star, 2015; Schneider, Rittle-Johnson, & Star, 2011), we argue that a skill that is more unconstrained in math would be evaluated by asking students to find *three different ways* to solve the problem: "What is 12×12?" Since there are more possibilities than three approaches (drawing a 12×12 square and counting cells, adding $12 + 12 + 12$ four times and then summing the totals, multiplying 10×12 to get 120 then adding $12 + 12$ to that sum, multiplying 24×6, etc.) this is an unconstrained task that calls upon student problem-solving skills and mathematical thinking.

Work from both language and literacy and math research would suggest that it is important for PreK programs to implement balanced instructional approaches that focus on supporting skills that can be conceptualized as *both* constrained and unconstrained. For example, while an emphasis in early literacy and math instruction on good practices (teaching phonological awareness, structured phonics, providing opportunities for reading fluency, developing knowledge of math facts, learning math procedures) is effective for improving early academic outcomes, success in early grade language and literacy or math does not automatically translate into long-term academic achievement, as shown by the fact that U.S. students score relatively well in Grade 4 international comparisons but much more poorly in Grades 8 and 10 (Provasnik et al., 2012). Second, a key determinant of success in post-primary reading comprehension is word

knowledge, a capacity indexed by (but not identical to) vocabulary (Kintsch, 1998).

Similarly, while procedural and conceptual knowledge in math are complementary and intertwined (Rittle-Johnson & Schneider, 2015), later math achievement requires a deep conceptual understanding of math as an abstract system, a capacity that goes beyond applying prescribed algorithms. Accordingly, successful math instruction initially prioritizes conceptual (or unconstrained) understanding over procedural (or constrained) skills (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998). Moreover, the idea of learning trajectories described by Sarama and Clements (2009) argues for the importance of students' mastery of constrained competencies as necessary for developing the capacity to engage in more unconstrained math activities. At the same time, associations between oral language in young children (including vocabulary as an index of word knowledge) and early reading skills (accurate decoding, fluency) are likely prerequisites for profiting from rich, topic-focused discussion and exposure to reading opportunities that build knowledge and reading comprehension (Beck & McKeown, 2001) as well as deep math understanding that goes beyond procedural training (Chapin & O'Connor, 2012).

PreK Instruction to Support Constrained and Unconstrained Skills

Although it may be ideal for PreK programs to use balanced instructional approaches that support students' constrained *and* unconstrained skills, work to date has shown that PreK programs demonstrate fairly low levels of instructional quality in general (Pianta, Downer, & Hamre, 2016) and learning opportunities for unconstrained skills may be limited. For example, evidence from The National Center for Early Development and Learning's Multi-State Study of Pre-Kindergarten and the Study of State-Wide Early Education Programs documented that PreK classrooms across 11 U.S. states spent 56% of time, on average, in learning activities including math, literacy, language, social science, science, art, and fine and gross motor skills (Early et al., 2010). The majority of this time was spent completing art projects, followed by literacy activities that targeted relatively *constrained* skills.

In contrast, PreK programs exhibiting higher levels of instructional quality—which we define as creating opportunities to acquire content knowledge and the language needed to conceptualize it, providing instruction across a range of skill domains (e.g., math, literacy, science), and using a range of high-quality practices to stimulate cognitive development (e.g., asking open-ended questions to build children's higher-order thinking skills, facilitating rich conversations)—may be more likely to *also* spend time supporting the development of *unconstrained* skills. Although constrained skills are critical to teach and serve as the foundation upon which more advanced skills can be developed (Snow & Matthews, 2016), they are often taught in isolation to support skill development that can be easily assessed and take up most instructional time in PreK classrooms (Claessens, Engel, & Curran, 2014; Engel, Claessens, Watts, & Farkas, 2016). Accordingly, children who do not attend PreK programs may quickly "catch up" to their peers when those constrained skills are taught in kindergarten (Bailey, Duncan, Odgers, & Yu, 2017).

Effects of PreK Programs on Children's Constrained and Unconstrained Skills

There is some experimental evidence to suggest that the short-term impacts of PreK programs may be larger on constrained versus unconstrained skills, perhaps reflecting a greater instructional focus on teaching constrained skills in early childhood (Lipsey et al., 2018). Wong et al. (2008) used an age-based regression discontinuity design to examine effects of public PreK programs across five states and found that program impacts on print awareness—a constrained skill—were four times larger than impacts on receptive vocabulary, an unconstrained skill. Yet, effects on unconstrained skills may be more likely to be sustained across time. For example, the National Head Start Impact Study collected the same five language and literacy assessments and one math assessment across the spring of PreK, kindergarten, and first grade for the 4-year-old cohort of study participants. Results revealed lasting (but small) impacts only on the Peabody Picture Vocabulary Test (PPVT) assessment of receptive vocabulary—the one measure of unconstrained vocabulary skills the team administered—in the spring of first grade (Puma et al., 2010).

It is also difficult to fully evaluate whether high-quality PreK programs are more likely to have lasting benefits on unconstrained versus constrained language and literacy and math skills simply because studies of early learning programs are more likely to *assess* constrained student outcomes. There is a wider variety of established, available

measures to assess constrained skills. In addition, these assessments are generally simple to collect because they can be done quickly after data collectors participate in a brief training (Snow & Matthews, 2016). For example, the PPVT assessment of receptive vocabulary—a measure of unconstrained skills—takes about 15 to 20 min to administer (Dunn & Dunn, 2007), relative to the Dynamic Indicators of Basic Early Literacy Skills assessment of constrained skills, which takes about 1 min per subscale to administer (Good et al., 2011).

Similarly, the Woodcock Johnson Applied Problems (WJAP) which assesses children's early math skills takes about 5–10 min to administer 4- and 5-year-old children. Although there is little work that has created separate measures to distinguish constrained from unconstrained math skills, the WJAP falls closer to the constrained end of this continuum for young children as it primarily assesses procedural skills like number recognition, counting, arithmetic (e.g., addition, subtraction, multiplication, and division), and shape recognition (Rittle-Johnson et al., 2015). For example, the start point item on this assessment for a 5-year-old asks the child to report how many dogs there are in a picture. The start point for a 6-year-old asks the child to use a picture of 3 cans to conduct a simple subtraction problem (3 cans minus 1 can) and the start point for a 7-year-old asks the child to look at a picture of 6 lollipops and subtract 1 (6 lollipops minus 1 lollipop). The assessment does not capture any information about *how* the child answered the question or solved the problem.

In contrast, we argue that the Research-based Early Math Assessment (Clements, Sarama, & Liu, 2008) is an example of a math measure that captures both constrained skills and a broader conceptual understanding of math skills—which we argue integrates more unconstrained skills—as the measure assesses algebraic thinking, composing shapes, patterning, comparing numbers, sequencing, and relational thinking in samples of young children. In contrast to the WJAP, the test has two components that assess numeracy and geometry. The starting point item for a 5-year-old on the geometry subtest shows the child an example of an isosceles triangle and asks if a picture of an obtuse triangle is also a triangle, following up to ask the child for his or her reasoning in answering this question. The starting point for a 6-year-old asks the child to essentially divide a pentagon into a trapezoid and triangle, while the starting point for a 7-year-old asks the child to look at five different pictures of angles and indicate which two angles are the same and explain why. A key part of this question is the focus on understanding how the child came to the correct answer and assessing his or her thought process. This short comparison serves to illustrate both why one could consider the Research-based Early Mathematics Assessment (REMA) as a more unconstrained measure and that the REMA is much more difficult and complex to use in assessing children, requiring a substantial amount of assessor training and certification. Given the differences in measures assessing language and literacy and math, there is a need for intentionally designed research that explicitly examines the effects of high-quality PreK programs on *both* constrained and unconstrained skills in these domains across time.

The Boston Public Schools Prekindergarten and Kindergarten Program

In order to test whether high-quality PreK has more lasting effects on unconstrained versus constrained skills, it is important to study a program that has demonstrated high levels of instructional quality and is using a curriculum that aims to support *both* types of skills. The current study addresses this need by leveraging data collected on students who attended the Boston Public Schools (BPS) prekindergarten program during the 2016–2017 academic year.

BPS's prekindergarten model consists of two evidence-based curricula: an adapted version of Opening the World of Learning (Schickedanz & Dickinson, 2004), a language and literacy curriculum that includes a social-emotional skills component in each unit, and Building Blocks (Clements & Sarama, 2007), an early mathematics curriculum that also promotes language development by requiring children to explain their mathematical reasoning verbally. The program is implemented in public schools, co-located with later elementary school grades, a structural feature that is theorized to be important for supporting high levels of program quality (Choi, Elicker, Christ, & Dobbs-Oates, 2016; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Play-based classroom instruction focuses on extending children's learning and deepening their understanding of language, literacy, and mathematical concepts through problem solving and peer interaction.

The district provides PreK teachers with some curriculum-specific training and in-class support from experienced early childhood coaches. Professional development to support model fidelity and improve teacher practice is the other core part of

the model. Teachers who are new to the curriculum are expected to attend 2 days of training prior to the start of the academic year. They can attend monthly professional development workshops hosted by the BPS Department of Early Childhood (DEC) staff on key topics relevant to implementing the curriculum. Teachers also have the option to participate in coaching sessions with a BPS DEC instructional coach (Weiland, McCormick, Mattera, Maier, & Morris, 2018). The model targets a number of constrained and unconstrained language and literacy and math skills, including vocabulary, early literacy, numeracy, geometry, conceptual thinking, problem-solving, executive functioning, and social-emotional skills. The curriculum was designed to promote teaching practices that enhance students' agency and thinking and help build communities of learners that are able to talk about their work and learn from one another. Recently collected observational data provide empirical evidence that the BPS prekindergarten model does balance instruction to support both constrained and unconstrained skills (Weiland et al., under review). We theorize that this stands in contrast to the broader set of formal PreK programs that prior work suggests focus more time on constrained than unconstrained skills (e.g., Early et al., 2010).

An earlier study using a regression discontinuity design found moderate to large effects of the program on constrained and unconstrained skills (language, literacy, and math) in the fall of kindergarten, and evidence of some larger subgroup effects (e.g., children from low-income and Latinx families; Weiland & Yoshikawa, 2013). The study further found that the instructional quality of BPS prekindergarten classrooms was much higher than the average quality of large-scale public prekindergarten programs nationally (Chaudry, Morrissey, Weiland, & Yoshikawa, 2017). However, this paper only assessed impacts at the start of kindergarten and there are questions about whether the PreK boost was sustained across time.

To address this need, Weiland et al. (2019) used an experimental lottery-based design to examine the impacts of the BPS prekindergarten program on state test scores, grade retention, and special education in third grade. Importantly, this analysis was limited to the sub-sample of students who applied to oversubscribed prekindergarten programs from 2007 to 2010, a group that was more advantaged than students who did not apply to oversubscribed programs. In addition, nearly all students in the control group in this study enrolled in some other type of center-based care as 4-year-olds, meaning that enrollment in BPS prekindergarten was compared to other center-based programming rather than to no PreK at all. Findings revealed no impacts of the BPS prekindergarten program on children's examined third grade outcomes, relative to the highly served comparison group. A complementary study by Weiland, Unterman, and Shapiro (in press) showed that most of this convergence effect occurred by the end of kindergarten. However, that study used the DIBELS only—a constrained measure of literacy skills—to assess outcomes and was unable to examine convergence in *unconstrained* skills.

Weiland et al. (2019) examination of long-term effects also considers four cohorts of students enrolled from 2007 to 2010. Yet, in 2012 BPS began implementing a kindergarten—second grade curriculum called *Focus on Early Learning* that implements content-rich curricula and professional development aligned with its prekindergarten model. The *Focus on Early Learning* program is designed to ensure that kindergarten teachers build effectively on what children are taught in PreK. Curricula across first and second grade are similarly aligned to support continuity in learning across later grades as well. A core part of the *Focus on Early Learning* program is the use of thematic units across grades. The thematic units are intended to help students make connections across content areas and develop their content-specific skills. The professional development supports in kindergarten are designed to look similar to those offered in prekindergarten.

Both children who do and do not attend the BPS prekindergarten program are thus exposed to *Focus on Early Learning* upon beginning formal kindergarten during their 5-year-old year (McCormick et al., 2020). Research conducted for a complementary study using the Individualizing Student Instruction observational measure (Connor et al., 2009) has shown that children in contemporary BPS prekindergarten and kindergarten classrooms are exposed to instruction supporting *both* constrained and unconstrained skills with *more* language and literacy instructional time spent on teaching unconstrained skills in prekindergarten (Weiland et al., under review). We know of no other preschool model in the United States that has demonstrated empirical evidence for balancing instruction to support both constrained and unconstrained skills. The BPS prekindergarten program (also described in this paper as the public PreK program) may thus be one of the best models available to study in order to examine the

constrained- unconstrained skills hypothesis for the PreK convergence phenomena. Data from complementary work suggest that classrooms implementing the public PreK model have higher levels of quality than community-based PreK options, including those implementing similar sets of curricula (Yudron, Weiland, & Sachs, 2016). Fidelity of implementation to the BPS prekindergarten model tends to be fairly high in public PreK programs and lower in community-based programs implementing a similar model (McCormick et al., 2020). As discussed in Guerrero-Rosada et al. (2020), the quality of the public PreK program experienced by students during the 2016–2017 year as measured using the instructional support domain on the Classroom Assessment Scoring System was generally higher than national averages.

But even with these experimental and descriptive data, there continues to be a need for further research that addresses extant gaps in prior work by: examining a representative sample; explicitly comparing students enrolled in BPS prekindergarten, students enrolled in other center-based care, and students not enrolled in center-based care at all; considering outcomes that differentiate constrained from unconstrained skills prior to third grade; and examining a more contemporary cohort of students who experienced the BPS prekindergarten program and the aligned *Focus on Early Learning* model in elementary school. In addition, given evidence that the benefits of high-quality PreK tend to be larger for low-income, DLL, and Hispanic students (Bloom & Weiland, 2015; Gormley, Gayer, Phillips, & Dawson, 2005; Lipsey et al., 2018), further examination of subgroup effects on constrained versus unconstrained skills over time is also warranted. Such tests can help reveal whether PreK programs are providing a compensatory mechanism to help disadvantaged children catch up to their peers on constrained skills in the short term, rather than closing gaps on the key unconstrained skills that stand to affect students' long-term academic trajectories.

To this end, the current study aims to build on prior work by using an exploratory approach to answer the following questions:

1. What is the association between attending the BPS prekindergarten program and children's constrained and unconstrained language and literacy skills in the fall and spring of kindergarten, relative to attending a non-public PreK program or not attending a formal PreK program?
2. What is the association between attending the BPS prekindergarten program and children's constrained and unconstrained math skills in the fall and spring of kindergarten, relative to attending a non-public PreK program or not attending a formal PreK program?
3. How do associations between enrollment in BPS prekindergarten and constrained and unconstrained skills in the fall and spring of kindergarten vary for lower-income, non-white, and DLL students?

Results will provide evidence for or against the unconstrained versus constrained skills hypothesis and help to better understand the PreK convergence puzzle currently challenging the field of early childhood education.

Method

Participants

The sample for the current study consists of 508 students attending the BPS kindergarten program during the 2017–2018 year. In the fall of 2016, we recruited 388 students from 41 public prekindergarten classrooms and 10 community-based (CBO) classrooms, nested within 20 public schools and 10 CBO centers to participate in this study. We were able to locate and collect data on 332 of those students after they transitioned into kindergarten with 14% of the study sample attriting from our sample. We recruited an additional 176 students in the fall of 2017 from 49 kindergarten classrooms in 21 public schools (the same kindergarten classrooms attended by our original prekindergarten sample). These 176 students had experienced a range of settings in the year prior to kindergarten—41% of them stayed home with a parent, family member or other adult, or attended a home-based daycare and the remaining 59% enrolled in a non-public PreK program. As discussed in more detail in the measures section below, we categorized the full sample of 508 kindergarten students into public PreK attenders ($N = 290$ students who attended the public BPS prekindergarten program during their 4-year-old year), non-public PreK attenders ($N = 147$ students who attended a CBO program, a private childcare center, or a Head Start center during their 4-year-old year), and non-PreK attenders ($N = 71$ students who enrolled in no formal PreK during their 4-year-old year).

Table 1
Sample Demographic Characteristics for Full Sample and Subgroups of Interest

	Full sample		Public PreK		Non-public PreK		No PreK	
Characteristic	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*
Student demographic characteristics								
Child age	5.60	0.30	5.61	0.29	5.57	0.31	5.61	0.33
Eligible for free/reduced price lunch	0.68	—	0.59	—	0.83	—	0.72	—
Female	0.51	—	0.49	—	0.54	—	0.50	—
Dual language learner	0.53	—	0.54	—	0.43	—	0.67	—
Asian	0.16	—	0.16	—	0.18	—	0.18	—
Black	0.28	—	0.20	—	0.46	—	0.20	—
Hispanic	0.32	—	0.32	—	0.25	—	0.46	—
Other race	0.03	—	0.05	—	0.00	—	0.03	—
White	0.21	—	0.27	—	0.11	—	0.13	—
Parent demographic characteristics								
Age of mother when first child was born	26.19	6.66	27.08	6.81	25.24	6.12	23.75	6.11
Total household size	4.25	1.36	4.25	1.22	4.04	1.59	4.66	1.47
At least one adult works 35 hr or more per week	0.90	—	0.93	—	0.86	—	0.80	—
Parent married	0.53	—	0.59	—	0.40	—	0.43	—
Parent age	28.70	8.62	29.11	8.27	28.75	9.10	26.47	9.32
Parent education: high school diploma/GED or less	0.31	—	0.30	—	0.29	—	0.48	—
Parent education: 2 year degree/technical or vocational certificatio	0.31	—	0.26	—	0.46	—	0.22	—
Parent education: 4 year degree	0.17	—	0.19	—	0.12	—	0.17	—
Parent education: more than a 4 year degree	0.21	—	0.25	—	0.13	—	0.13	—

Note. Full sample of 508 kindergarten students is grouped into: public Prek attenders ($N = 290$), non-public PreK attenders ($N = 147$), and non-PreK attenders ($N = 71$). GED = general education diploma.

The demographic characteristics of the study sample are presented in Table 1. As illustrated there, the majority of the sample was eligible for free or reduced price lunch (FRPL) and the students were diverse with respect to racial/ethnic background and parental education, among other characteristics. As described more fully in Supporting Information Appendix B, the public schools in the sample are representative of the broader population of BPS elementary schools. The students in the analytic sample are generally representative of the broader group of students who enrolled in the study in preK on eligibility for FRPL, race (Black, White, Hispanic, mixed, or other race), DLL status, and family or parent characteristics. However, the 56 students we could not follow into kindergarten were more likely to have attended prekindergarten in a CBO (63% of attrited sample attended a CBO versus 37% of attrited students who attended a public school). The students who originally enrolled in the study and the 176 students who were recruited into the study in Fall 2017 were representative of the broader population of students enrolled in kindergarten in the district.

Procedure

School and Classroom Recruitment

In 2016, before the start of the prekindergarten year, we randomly selected 25 public schools to participate in the study from the 76 schools in the district offering the public prekindergarten program. Twenty-one agreed. We used one school as a pilot school for developing new measures and the remaining 20 schools made up the public school sample in the first year of the study. We randomly selected 10 of the 11 CBOs in Boston implementing the BPS prekindergarten model to participate in the study and they all agreed. These CBOs were connected to the BPS DEC and were receiving training and coaching to implement the BPS curricular model during the 2016–2017 year. The experiences of students in these CBOS are likely qualitatively different from students enrolled in the public PreK classrooms because teachers were only in the early implementation phases of the model during the 2016–2017 year and the centers continued to operate outside the purview of the broader public school system. Further, the teachers in CBO centers

were not subject to the same educational requirements (master's degree within 5 years) as public school-based teachers, nor where they paid on the same scale. It is important to note that children who attended one of the CBOs implementing the BPS prekindergarten model are considered non-public PreK attenders in this study. However, the full group of non-public PreK attenders consists of these students enrolled from the CBOs, as well as students who enrolled in the study in the fall of kindergarten and attended a private childcare center during their 4-year-old year and students who attended a Head Start program during their 4-year-old year. Thus, as discussed in more detail below, we group students enrolled from CBOs together with children who attended private childcare programs and Head Start centers during their 4-year-old year.

We asked all prekindergarten teachers assigned to general education or inclusion classrooms in each of the 20 public schools to participate in the study in the fall of 2016. We randomly selected one classroom serving 4-year-old students within each CBO to participate. Ninety-six percent ($N = 51$) of teachers across public schools ($N = 20$) and CBOs ($N = 10$) agreed. We followed sample children into kindergarten across 52 schools and 98 kindergarten classrooms. We asked children's kindergarten teachers to participate and 95% agreed.

Student Recruitment

We attempted to collect active consent for all prekindergarten students enrolled in participating classrooms. Research staff sent home backpack mail providing an overview of the study and a consent form for the parent to complete and return. We did regular sweeps to pick up consents. Eighty-one percent of all children in participating classrooms had parent consent to participate in the study. Of the children with parent consent, the team randomly selected 50% (~6–10 per classroom) to participate in student-level data collection activities. We repeated this process in the kindergarten classrooms participating in the study in the fall of 2017 and enrolled 78% of students in the participating classrooms who had not attended the public school BPS prekindergarten program in the 2016–2017 academic year. This group included children who attended a CBO implementing the BPS model during their 4-year-old year, as well as children who attended a private childcare center or Head Start center, or did not experience any formal center-based PreK during their 4-year-old year.

Direct Assessments

We trained research staff to reliability and then collected direct assessments of academic skills in the fall of 2017 (September 22 through December 7) and spring of 2018 (April 1 through June 15), when study students were enrolled in kindergarten. We collected assessments for 75% of the student sample by October 22 (within the first 5 weeks of school). In line with a wealth of studies examining the short-term benefits of PreK (e.g., Gormley et al., 2005; Hustedt, Barnett, Jung, & Friedman, 2010; Ludwig & Miller, 2007; Weiland & Yoshikawa, 2013; Wong, 2008), assessing outcomes in the fall of the kindergarten year—prior to children being exposed to the bulk of kindergarten instruction—is a standard practice for estimating the effect of PreK enrollment on children's outcomes. We used the Pre-language Assessment Scale (preLAS; Duncan & DeAvila, 1998) Simon Says and Art Show tests to determine the administration language for a subset of assessments (Barrueco, Lopez, Ong, & Lozano, 2012). The preLAS assesses pre-literacy skills and an individual's proficiency in English. Of the 508 children in the study sample, 15 did not pass the preLAS and completed a subset of assessments in Spanish in the fall and 4 students did not pass the preLAS and completed assessments in Spanish in the spring.

Administrative Data From the School District

We accessed administrative records from the BPS district on students' demographic characteristics, history of enrollment in the BPS prekindergarten program, classroom and school membership at the end of the kindergarten year, and teacher-collected assessments of student literacy skills in the fall and spring of kindergarten.

Parent Survey

In the fall of 2016 and 2017, we contacted the consenting parents of all students who were selected for the study sample to complete a 20-min survey via text message and email. Parents received biweekly reminders to complete the survey. We translated the surveys into Spanish, Vietnamese, and Mandarin. Parents provided demographic information about themselves and their child. Across both survey waves, 86% of respondents were mothers and 12% were fathers. All parents received a $25 gift card for completing the survey. In total, 95% of the parents included in the current

study completed a parent survey in 2016 or 2017, allowing for fairly complete parent-reported covariate data.

Measures

Experience in Early Childhood Education

In the fall of 2017, we asked parents to report where their child spent the majority of his or her time during a regular week during the 2016–-2017 year. We then coded these responses to describe whether children were primarily enrolled in the BPS prekindergarten program (described in the methods and results sections as public PreK), a non-public PreK program (e.g., Head Start, private child care center, CBO PreK program), or did not attend formal PreK (e.g., stayed at home with a parent, family member, friend, and/or attended a home-based childcare). Within the study sample, there are 147 children enrolled in the non-public PreK group. Forty-seven percent of the students in this group attended a CBO implementing the BPS prekindergarten model during their 4-year-old year, 24% attended a private child care center during their 4-year-old year, and 29% attended a Head Start program during their 4-year-old year. We confirmed these categorizations through on-line searches of the reported locations coupled with follow-up phone calls to centers. We cross-referenced codes with district data and used district data from the end of kindergarten to further clean this variable, describing a child as being a public PreK attender if he or she had been enrolled for 89 days or more (at least half the year) in the public PreK program. We then dummy coded these variables by assigning a 1 to a public PreK attender and a 0 otherwise, and a 1 to a non-public PreK attender and a 0 otherwise (with non-PreK attender as the reference group). We initially considered treating students enrolled in CBO PreK programs as a separate group from students in private childcare and Head Start centers. However, examination of descriptive statistics suggested that the skills of these children in the fall and spring of kindergarten were similar and treating them separately created small samples that might have lacked adequate statistical power to detect group-based differences. Moreover, as discussed earlier, the structural characteristics of these PreK contexts were more similar to one another than they were to the public PreK program. Even so, we conduct a number of robustness checks (described in more detail below) to examine how results might have differed if we had not grouped these students together.

Unconstrained Language Skills

We used the PPVT, 4th ed. (Dunn & Dunn, 2007) to directly assess children's receptive language skills in the fall and spring of kindergarten. We used this assessment as a measure of children's unconstrained language skills. The PPVT IV is a nationally normed measure that has been used widely in diverse samples of young children. The test has excellent split-half and test–retest reliability estimates, as well as strong qualitative and quantitative validity properties (Dunn & Dunn, 2007). It requires children to choose (verbally or nonverbally) which of four pictures best represents a stimulus word. We used the PPVT raw score as our outcome measure. We assessed all children on the PPVT—regardless of whether they passed the PreLAS language screener—in order to describe an equivalent measure of receptive language skills in English across the sample. The proportion of students who did not pass the preLAS was fairly small (< 3% of the sample).

Constrained Literacy Skills

We used subtests from the teacher-reported Dynamic Indicators of Basic Literacy Skills—Next (DIBELS; Good et al., 2011) to measure children's constrained literacy skills. Administered subtests measured children's letter knowledge (Letter Naming Fluency; LNF), phonological awareness (First Sound Fluency and Phoneme Segmentation Fluency; FSF and PSF), and alphabetic principle (e.g., letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds; Nonsense Word Letter Sounds and Nonsense Words Whole Word Read: NWLS and NWWWR). These subtests have good reliability and good concurrent, predictive, and discriminant validity properties, are widely used, and are sensitive to intervention effects (e.g., Biancarosa, Bryk, & Dexter, 2010). BPS teachers administered the FSF and LNF subtests in the fall of kindergarten and the LNF, PSF, NWLS, and NWWWR in the spring of kindergarten spring. It is important to note that BPS teachers administered different subtests at different time points during the kindergarten year as DIBELS is designed to follow children's developmental progression in literacy (e.g., phoneme segmentation is not measured until the end of kindergarten, when it is developmentally and instructionally expected for most students; Good et al., 2011). Although the subtests used in this study do differ across the year, the DIBELS

measure as a whole captures the underlying construct of constrained literacy skills.

Unconstrained Math Skills

We used the REMA (Clements et al., 2008) to assess children's unconstrained math skills in the fall and spring of kindergarten. We recognize that the REMA does include some constrained items but argue that it falls more on the unconstrained end of the constrained—unconstrained continuum. The REMA is a hands-on, one-on-one assessment that measures core mathematical abilities of children ages 3–8. As described further in Sarama et al. (2012), children's abilities are assessed on the REMA according to theoretically and empirically based developmental progressions that underlie learning trajectories. The assessment captures the following skills within the subdomain of numerical progressions: verbal counting, object counting, subitizing, number comparison, number sequencing, connection of numerals to quantities, number composition and decomposition, adding and subtracting, and place value. Geometry progressions include shape recognition, congruence, construction of shapes, and spatial imagery, as well as geometric measurement and patterning. Work on the REMA and learning trajectories differentiating procedural and conceptual math skills (Sarama & Clements, 2009) would conceptualize such competencies as number sequencing, number composition and decomposition, construction of shapes, spatial imagery, and measurement and patterning as being more unconstrained while competencies like counting, addition, subtraction, shape recognition, and congruence would be more constrained. Importantly, assessors code not only whether the child answered each item correctly or not, but also the difficulty of the strategy that the child used to respond to the item. Students receive more points on the assessment if they use more advanced or complex strategies to answer items. All items—including the strategy codes—are ordered by Rasch item difficulty and children stop the assessment after making three consecutive errors. All assessors administering the REMA in the current study had to achieve 90% or higher on a mock administration. Staff who worked with the test developer certified assessors. The alpha reliabilities for the test total scores (referred to as t scores in this manuscript) have been shown to range from .92 to .94 (Sarama et al., 2012) with prior studies also demonstrating evidence of construct and concurrent validity (Clements et al., 2008).

Constrained Math Skills

We used the WJAP III (Woodcock, McGrew, & Mather, 2001) subtest to assess children's constrained math skills in the fall and spring of kindergarten. We assessed Spanish-speaking children who did not pass the PreLAS language screener using the equivalent Spanish language version of the assessment from the Batería III Woodcock Muñoz (Schrank, McGrew, Ruef, & Alvarado, 2005). The WJ/WM Applied Problems assessment is a numeracy and early mathematics measure that requires children to perform calculations to analyze and solve arithmetic problems. Its estimated test–retest reliability for 2- to 7-year-old children is .90 (Woodcock et al., 2001) and it has been used with diverse populations (Gormley et al., 2005; Wong et al., 2008). We present results using the raw score of the measure. We combined scores from the English and Spanish assessments together.

Child Characteristics From Administrative Data

Using administrative data, we created a series of indicators to describe children's race/ethnicity (Black, Hispanic, Asian, or Other Race/Ethnicity (including mixed race children)), coding 1 if the child fell into the indicated category and 0 otherwise (reference group White). We used similar indicators to describe eligibility for FRPL (1 if eligible; 0 if not) and gender (1 = female; 0 = not female). We set a dummy variable for DLL equal to 1 if the parent reported that there was a language other than English spoken at home and 0 otherwise. We used the child's birthdate to calculate child age on September 1, 2017. These covariates have been shown to predict children's outcomes across studies (Choi, Jeon, & Lippard, 2018; Reardon & Portilla, 2016).

Family Characteristics From Parent Survey

Parents reported on demographic characteristics in the fall of the PreK year and we used these indicators as covariates. These variables indicated whether there was at least one parent in the home working full-time (35 hr/week or more) and whether the parent was married or lived with a partner. We used continuous variables to describe the age of the child's mother at her first birth, the number of people living in the household, and the parent respondent's age in the fall of the prekindergarten year. We then included three dummy variables to describe the reporting parent's level of education as a general proxy for socioeconomic

status—high school diploma or general education diploma (GED) or less, some college or 2-year degree, and 4-year degree (with graduate work or graduate degree as the reference group). These covariates have been shown to predict children's outcomes (Choi et al., 2018; Powell, Son, File, & San Juan, 2010; Reardon & Portilla, 2016) across studies.

Analytic Approach

Missing Data

Overall, there was a relatively low amount of missing data. All students had complete data on child-level information provided by the school district. About 13% of students were missing parent survey covariate data, 10% were missing some outcome data on one of the assessments our research team collected, and 24% of students were missing information on the DIBELS outcome that was collected by the district. We used multiple imputation to handle missing covariate data and then compared results with complete case analysis and imputation of both predictors and outcomes. Results were consistent across specifications. See Supporting Information Appendix A for more details on our multiple imputation approach and results.

Descriptive Analysis

To describe the sample before examining our research questions, we computed descriptive statistics on demographic variables and assessment scores for students who attended public PreK, nonpublic PreK, and those who did not attend formal PreK. As part of the descriptive analysis, we also calculated the zero-order correlations between the binary indicators for PreK attender groups and each of the outcomes in the fall and spring of kindergarten. Findings from these correlations help demonstrate whether measures of constrained-unconstrained language and literacy and math skills were empirically distinct from one another.

Multi-Level Modeling to Examine Study Research Questions

We then used multi-level modeling to examine associations between enrollment in public PreK and students' skills in kindergarten, and enrollment in non-public PreK and skills in kindergarten. Because students in our sample were nested within classrooms nested within schools in kindergarten, we fit null models for the spring assessment scores to calculate intraclass correlations (ICCs) and examine the extent to which observations were non-independent at these levels. The ICCs representing variation between classrooms in the outcomes ranged from .03 through .31 while the ICCs representing variation between schools ranged from .03 to .15. Given that the variation at both the classroom- and school-level was non-negligible, we used a three-level model with random intercepts for classrooms and schools to examine all of our research questions. However, we also fit these models using fixed intercepts for schools and classrooms and using regression adjusted inverse probability weighting (also known as double robust estimation) as robustness checks to examine the sensitivity of our results to model specification.

To answer research questions 1 and 2, we regressed each outcome in the fall and spring of kindergarten on a dummy variable for enrollment in Public PreK (i.e., the BPS prekindergarten program) and a dummy variable for enrollment in non-public PreK. We then added covariates to the models in two conceptual blocks (first block = child-level covariates; second block = parent-level covariates). We examined the stability of the point estimates for enrollment in public and non-public PreK across the models to examine how sensitive parameter estimates were to covariates. The base equation for these models is as follows:

$$Y_{ijk} = \beta_0 + \beta_1 \text{PublicPreK}_{ijk} + \beta_2 \text{NonpublicPreK}_{ijk} + \gamma_{ijk} + \mu_{jk} + \zeta_k + \varepsilon_{ijk}, \quad (1)$$

where i denotes students and j and k reesent kindergarten classrooms and schools, respectively, Y is the child-level outcome measure, *PublicPrek* is a dichotomous indicator set to one if student i attended the BPS PreK program and 0 otherwise, and *NonpublicPreK* is a dichotomous indicator set to one if student i attended PreK in a non-public setting and 0 otherwise. γ is a vector of child- and family level covariates, μ is the classroom-level random intercept, ζ is a school-level random intercept, and ϵ is the child-level error term. The coefficients on *PublicPreK* and *NonpublicPreK* and their associated p-values are the parameters of interest. We then used a Wald test to directly compare the coefficients on *PublicPreK* and *NonpublicPreK* and determine how these PreK attender groups compared to one another (Cohen, Cohen, Aiken, & West, 2013). This difference (and its associated p-value) indicates

whether the outcomes for public PreK attenders are statistically significant different from the outcomes for non-public PreK attenders. We calculated standardized associations by dividing the parameter estimates by the standard deviation of the outcome (Shadish, Cook, & Campbell, 2002). We used a similar model to answer research question 3, but also included interactions between PreK attender groups and subgroups of interest in the model. See Supporting Information Appendix C for full model.

Results

Descriptive Analysis

We found substantial differences in the demographic characteristics of children in our public PreK, non-public PreK, and non-PreK attender groups. As illustrated in more detail in Table 1, public PreK attenders were significantly less likely than the other two groups to be eligible for free lunch, more likely to be white, more likely to have parents who were married, and more likely to have a parent with more than a 4-year degree. Relative to public PreK attenders and non-attenders, non-public PreK attenders were more likely to be Black and more likely to have a parent with a 2-year degree or technical or vocational certification. Relative to public and non-public PreK attenders, non-attenders were more likely to be Dual Language Learners and Hispanic, and more likely to have a parent with only a high school diploma or GED or less. Table 2 then provides descriptive statistics on students' constrained and unconstrained language and literacy and math skills in the fall and spring of the kindergarten year, broken out by PreK attender group. As illustrated there, public PreK attenders generally outperformed non-public PreK attenders and non-attenders on assessments of constrained and unconstrained skills in the fall of the kindergarten year. Differences—as explored in more detail in our covariate-adjusted models discussed below—appeared to shrink across the year for constrained literacy skills in particular.

Correlational analyses revealed that our measures of constrained and unconstrained language and literacy in the fall and spring of the kindergarten year were empirically distinct from one another. However, we did find larger correlations between our measures of constrained and unconstrained math skills (see Supporting Information (S1) for correlation matrix). For example, the PPVT and the DIBELs LNF subtest were correlated at $r = .39$ ($p < .01$) and $r = .43$ ($p < .01$) in the fall and spring, respectively. The WJAP and REMA were correlated at $r = .76$ ($p < .01$) in the fall and $r = .77$ ($p < .01$) in the spring of kindergarten.

Table 2
Descriptive Statistics for Constrained and Unconstrained Language/Literacy and Math Skills in Kindergarten

	Fall of kindergarten		Spring of kindergarten	
Variable of interest	*M*	*SD*	*M*	*SD*
PPVT raw score				
Full sample	87.57	27.70	101.28	26.81
Public PreK	94.46	26.08	107.24	25.38
Non-public PreK	81.79	24.86	95.79	25.48
No PreK	70.58	30.37	85.50	26.78
Woodcock Johnson Applied Problems raw score				
Full sample	15.83	5.27	19.28	4.65
Public PreK	17.15	5.09	20.15	4.68
Non-public PreK	14.82	4.70	18.15	4.12
No PreK	12.43	5.23	17.50	4.49
REMA *t* score				
Full sample	40.03	5.74	43.53	5.11
Public PreK	41.30	5.49	44.46	5.18
Non-public PreK	38.44	5.58	42.36	4.47
No PreK	37.55	5.54	41.55	4.92
DIBELS Letter Naming Fluency score				
Full sample	23.50	17.08	50.38	17.98
Public PreK	27.01	16.80	50.89	18.66
Non-public PreK	20.95	15.87	51.64	16.31
No PreK	13.50	15.79	45.63	17.59
DIBELS First Sound Fluency score				
Full sample	15.24	12.74	—	—
Public PreK	18.60	12.67	—	—
Non-public PreK	11.07	11.48	—	—
No PreK	8.81	10.46	—	—
DIBELS Phoneme Segmentation Fluency score				
Full sample	—	—	41.01	18.22
Public PreK	—	—	43.29	17.16
Non-public PreK	—	—	39.65	19.54
No PreK	—	—	33.48	18.14
DIBELS nonsense word fluency correct letter sounds score				
Full sample	—	—	38.57	26.15
Public PreK	—	—	40.41	28.09
Non-public PreK	—	—	37.72	23.82
No PreK	—	—	32.05	20.09
DIBELS nonsense word fluency whole word read score				
Full sample	—	—	6.91	10.92
Public PreK	—	—	8.36	11.52
Non-public PreK	—	—	5.34	10.34
No PreK	—	—	3.54	7.80

Note. Full sample of 508 kindergarten students is grouped into: public Prek attenders ($N = 290$), non-public PreK attenders ($N = 147$), and non-PreK attenders ($N = 71$). PPVT = Peabody Picture Vocabulary Test; REMA = Research-based Early Mathematics Assessment; DIBELS = Dynamic Indicators of Basic Literacy Skills.

Research Question 1: What Is the Association Between Attending the BPS Prekindergarten Program and Children's Constrained and Unconstrained Language and Literacy Skills in the Fall and Spring of Kindergarten, Relative to Attending a Non-Public PreK Program or Not Attending a Formal PreK Program?

We found that results were fairly consistent across models, even after adjusting for conceptual blocks of covariates. As such, we present the results from the fully controlled models throughout this section (Model 3 in Tables 3 and 4; Model 1 presents results with no controls and Model 2 presents results with student-level covariates from administrative data). As illustrated in Table 3, we found that the associations between enrollment in public PreK and unconstrained language skills were statistically significant in both the fall ($\gamma = 15.79$, $SE = 3.06$, $p < .001$; std. association = .57) and the spring of kindergarten ($\gamma = 12.79$, $SE = 3.06$, $p < .001$; std. association = .48). Associations between enrollment in non-public PreK and unconstrained language skills were also statistically significant in the fall ($\gamma = 10.07$, $SE = 3.37$, $p < .01$; std. association = .36) and spring of kindergarten ($\gamma = 8.05$, $SE = 3.44$, $p < .05$; std. association = .30). Findings from a Wald test revealed that the magnitude of these effects was larger for public PreK relative to non-public PreK in the fall (mean difference = 5.72, $SE = 2.48$, $p < .05$). The magnitude of the effect for public PreK was also larger than non-public PreK in the spring but the difference was not statistically significant (mean difference = 4.74, $SE = 2.60$, $p = .07$).

The associations between public PreK and the DIBELS LNF subtest were statistically significant in the fall of kindergarten ($\gamma = 10.70$, $SE = 2.33$, $p < .001$; std. association = .63; see Table 3), but were not sustained into the spring of kindergarten ($\gamma = 2.26$, $SE = 2.61$, $p = .39$; std. association = .13; see Table 3). We found a similar trend for the associations between enrollment in non-public PreK and the DIBELS LNF subtest in the fall ($\gamma = 7.13$, $SE = 2.64$, $p < .05$; std. association = .42) and spring of kindergarten ($\gamma = 4.66$, $SE = 2.90$, $p = .11$; std. association = .26). The magnitude of the association in the fall of kindergarten was larger for public PreK attenders relative to non-public PreK attenders but was not statistically significant ($\gamma = 3.58$, $SE = 1.98$, $p = .07$). The difference in these associations between public and non-public PreK attenders was not statistically significant in the spring ($\gamma = -2.00$, $SE = 2.02$, $p = .58$). Although there was a statistically significant association between enrollment in public PreK and the DIBELS FSF subtest in the fall of kindergarten ($\gamma = 6.95$, $SE = 1.67$, $p < .001$; std. association = .55), this association was non-significant for non-public PreK attenders. A Wald test revealed that the association was also larger for public PreK attenders compared to non-public PreK attenders (mean difference = 4.94, $SE = 1.12$, $p < .05$). There was a statistically significant association between enrollment in public PreK and the DIBELS PSF subtest in the spring of kindergarten ($\gamma = 5.30$, $SE = 2.26$, $p < .05$; std. association = .29), but no statistically significant associations between public PreK and either of the DIBELS word fluency outcomes in the spring of kindergarten. There were no associations between enrollment in non-public PreK and any of the DIBELS outcomes measured in the spring of kindergarten. Finally, there were no statistically significant differences between public PreK and non-public PreK attenders on the DIBELS PSF, NWLS, and NWWWS subtests in the spring of kindergarten. Importantly, as explained earlier, we were able to examine letter fluency in the fall and spring of kindergarten, but could only assess FSF in the fall of kindergarten, and then phoneme segmentation and word fluency at the end of kindergarten, following the developmental continuum for literacy in these years (Good et al., 2011). Although only one of the subtests remained consistent across time, the DIBELS subtests taken together capture the underlying construct of constrained literacy skills.

Research Question 2: What Is the Association Between Attending the BPS Prekindergarten Program and Children's Constrained and Unconstrained Math Skills in the Fall and Spring of Kindergarten, Relative to Attending a Non-Public PreK Program or Not Attending a Formal PreK Program?

In the next set of models examining math outcomes (see Table 4), we found that the associations between public PreK and unconstrained math skills scores were statistically significant in both the fall ($\gamma = 2.30$, $SE = 0.67$, $p < .001$; std. association = .40) and the spring of kindergarten ($\gamma = 1.52$, $SE = 0.60$, $p < .05$; std. association = .30). However, the associations between non-public PreK and unconstrained math skills were non-significant in the fall and spring of kindergarten. Results from Wald tests revealed that associations were larger for public PreK attenders relative to non-public PreK attenders in both the fall (mean difference = 1.60, $SE = 0.45$, $p < .05$) and the spring (mean difference = 1.00, $SE = 0.39$, $p < .05$). The association between public PreK and constrained math skills was statistically significant in the fall ($\gamma = 3.44$, $SE = 0.60$,

Table 3
Multi-Level Models Examining Associations between Enrollment in PreK and Language/Literacy Skills with Multiple Imputation of Covariates

	Fall of kindergarten							Spring of kindergarten						
	Model 1		Model 2		Model 3			Model 1		Model 2		Model 3		
Fixed effect	γ	*SE*	γ	*SE*	γ	*SE*	Std. association	γ	*SE*	γ	*SE*	γ	*SE*	Std. association
PPVT raw score														
Intercept	74.93	3.63*	4.98	18.57	−9.19	20.31	—	90.30	3.56*	30.44	19.38	19.32	20.90	—
Public PreK	17.61	3.47*	17.43	3.03*	15.79	3.06*	0.57	14.89	3.42*	14.91	3.03*	12.79	3.06*	0.48
Non-public PreK	9.55	3.76*	11.38	3.35*	10.07	3.37**	0.36	8.66	3.79***	9.61	3.43***	8.05	3.44***	0.30
DIBELS Letter Naming Fluency score														
Intercept	14.07	2.32*	−34.04	14.20***	−43.71	15.47**	—	46.57	2.69*	16.39	15.45	6.07	16.86	—
Public PreK	12.82	2.46*	11.80	2.31*	10.70	2.33*	0.63	4.28	2.65	3.63	2.58	2.26	2.61	0.13
Non-public PreK	6.51	2.71***	8.01	2.60**	7.13	2.64**	0.42	3.95	2.91	5.54	2.86	4.66	2.90	0.26
DIBELS First Sound Fluency score														
Intercept	10.26	1.75*	−14.66	10.17	−22.81	10.91***	—	—	—	—	—	—	—	—
Public PreK	7.58	1.78*	7.72	1.69*	6.95	1.67*	0.55	—	—	—	—	—	—	—
Non-public PreK	1.37	1.95	2.88	1.89	2.01	1.88	0.16	—	—	—	—	—	—	—
DIBELS Phoneme Segmentation Fluency score														
Intercept	—	—	—	—	—	—	—	37.86	2.62*	32.98	13.16***	24.49	14.35	—
Public PreK		—	—	—	—	—	—	6.56	2.24**	6.47	2.23**	5.30	2.26***	0.29
Non-public PreK	—	—	—	—	—	—	—	4.32	2.45	5.41	2.47***	4.57	2.50	0.25
DIBELS nonsense word fluency correct letter sounds score														
Intercept	—	—	—	—	—	—	—	32.99	3.79*	−27.46	22.51	−42.38	24.20	—
Public PreK	—	—	—	—	—	—	—	6.61	3.90	4.88	3.73	2.59	3.76	0.10
Non-public PreK	—	—	—	—	—	—	—	4.48	4.29	6.25	4.16	4.92	4.18	0.19
DIBELS nonsense word fluency whole word read score														
Intercept	—	—	—	—	—	—	—	4.13	1.55*	−15.40	9.40	−20.29	10.35	—
Public PreK	—	—	—	—	—	—	—	3.88	1.62***	3.07	1.56***	2.25	1.58	0.21
Non-public PreK	—	—	—	—	—	—	—	1.50	1.77	2.37	1.73	1.94	1.75	0.18
Covariates														
Child level covariates				X		X					X		X	
Family level covariates						X							X	

Note. PPVT = Peabody Picture Vocabulary Test; DIBELS = Dynamic Indicators of Basic Literacy Skills.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 4
Multi-Level Models Examining Associations Between Enrollment in PreK and Math Skills With Multiple Imputation of Covariates

Fixed effect	Fall of kindergarten							Spring of kindergarten						
	Model 1		Model 2		Model 3			Model 1		Model 2		Model 3		
	γ	*SE*	γ	*SE*	γ	*SE*	Std. association	γ	*SE*	γ	*SE*	γ	*SE*	Std. association
REMA *t* score														
Intercept	37.69	0.74***	14.60	4.33***	11.52	4.65*	—	41.58	0.70***	28.27	3.86***	24.05	4.12***	—
Public PreK	3.33	0.73***	2.85	0.68***	2.30	0.67***	0.40	2.45	0.66***	2.04	0.60***	1.52	0.60*	0.30
Non-public PreK	0.73	0.82	0.99	0.76	0.70	0.76	0.12	0.55	0.75	0.80	0.69	0.52	0.68	0.10
Woodcock Johnson Applied Problems raw score														
Intercept	12.59	0.69***	−4.12	3.69	−6.49	4.02	—	17.39	0.61***	2.90	3.58	0.01	3.87	—
Public PreK	4.07	0.67***	3.65	0.60***	3.44	0.60***	0.65	2.43	0.61***	1.91	0.55***	1.60	0.56**	0.31
Non-public PreK	2.20	0.73**	2.67	0.66***	2.60	0.66***	0.49	0.50	0.69	0.65	0.63	0.49	0.63	0.11
Covariates														
Child level covariates				X		X					X		X	
Family level covariates						X							X	

Note. REMA = Research-based Early Mathematics Assessment.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

$p < .001$; std. association = .65) and spring of kindergarten (γ = 1.60, *SE* = 0.56, $p < .01$; std. association = .31). The association between non-public PreK and constrained math skills was statistically significant in the fall of kindergarten (γ = 2.60, *SE* = 0.66, $p < .01$; std. association = .49), and not sustained into the spring of kindergarten (γ = .49, *SE* = 0.63, $p = .65$; std. association = .11). Results from Wald tests revealed that the magnitude of these associations was larger for public PreK attenders relative to non-public PreK attenders in both the fall (mean difference = .84, *SE* = 0.41, $p < .05$) and the spring (mean difference = 1.11, *SE* = 0.52, $p < .05$).

Research Question 3: How Do Associations Between Enrollment in BPS Prekindergarten and Constrained and Unconstrained Skills in the Fall and Spring of Kindergarten Vary for Lower-Income, Non-White, and DLL Students?

Results from the models examining variation in associations by family income, race/ethnicity, and DLL status revealed statistically significant interactions between public PreK and non-public PreK and the dummy for low income in predicting both the constrained and unconstrained math skills in the fall of kindergarten. The association between enrollment in public PreK and children's constrained and unconstrained math skills at the start of kindergarten, relative to not attending PreK at all, was larger for higher-income children (WJAP: γ = −2.86, *SE* = 1.24, $p < .05$; REMA: γ = −3.81, *SE* = 1.41, $p < .01$). In addition, the association between enrollment in non-public PreK and children's constrained and unconstrained math skills, relative to not attending PreK at all, was larger for higher-income children (WJAP: γ = −3.03, *SE* = 1.44, $p < .05$; REMA: γ = −3.66, *SE* = 1.65, $p < .05$). See Supporting Information Appendix E Figures E1 and E2 for an illustration of these interactions. The interactions were not significant at the end of kindergarten suggesting that these group-based differences were not sustained into the spring of kindergarten. The association between enrollment in public PreK and children's unconstrained language skills in the fall of kindergarten was larger for DLLs relative to non-DLLs (γ = 13.66, *SE* = 6.31, $p < .05$). Associations did not vary for children who attended non-public PreK. See Supporting Information Appendix E Figure E3 for an illustration.

Robustness Checks

We conducted a number of robustness checks to examine the sensitivity of our results to modeling

decisions and potential concerns about the sample and data, including approaches for handling missing data, alternative model specifications (fixed effects regressions and adjusted inverse probability of treatment weighting, Guo & Fraser, 2015), and selection by child characteristics into kindergarten classrooms of varying quality. As a further robustness check, we refit our models for the first two research questions by considering different groupings of PreK attenders to determine whether results varied considerably when the difference between the public and non-public PreK groups was more clean-cut (using four indicators to represent PreK attendance—public PreK, CBO PreK implementing BPS model, private childcare, and Head Start). We found that results were robust across these alternative approaches and specifications (see details in Supporting Information Appendix A).

Discussion

Our findings provide evidence that the unconstrained versus constrained skills hypothesis may help explain the PreK fadeout-convergence phenomenon. We found that associations between enrollment in the BPS prekindergarten program and students' outcomes were more strongly sustained through the spring of kindergarten for unconstrained language and math skills, compared to constrained literacy and math skills. There were sustained benefits of non-public PreK on unconstrained language skills through the spring of kindergarten, but not on unconstrained math skills or either of the constrained outcome domains. This is one of the first efforts to our knowledge to explicitly test whether there are more likely to be sustained associations between PreK and unconstrained skills compared to constrained skills. By leveraging data from the BPS prekindergarten and kindergarten programs, the study was uniquely situated to test this theory. The public PreK program of interest explicitly focused on supporting *both* constrained and unconstrained skills through a curriculum and professional development model implementing theme-based, rich content to support student background knowledge and critical thinking; project-based work; play-based, intentional instruction; small group and student-directed activities (rather than a predominance of whole-group activities); differentiation of instruction to limit repetition of content students have already learned; and training, coaching, and monthly curriculum-focused seminars for teachers. Furthermore, the kindergarten program was designed to build on the skills acquired in PreK rather than repeating PreK material.

Exposure to PreK Instruction Supporting Constrained and Unconstrained Skills

There were small inklings from prior work (e.g., Puma et al., 2010) that the effects of public PreK on unconstrained language outcomes might be more lasting than on constrained literacy skills. Even so, the large majority of studies conducted in the last 15 years have not shown evidence that effects on unconstrained skills were more likely than constrained skills to be sustained across time (e.g., Hill, Gormley, & Adelstein, 2015; Lipsey et al., 2018). However, there is also limited evidence about the extent to which the other PreK models that have been evaluated have integrated instructional supports for children's unconstrained and constrained skills and few other studies have been able to assess both constrained and unconstrained skills as outcomes. Complementary data suggest that children in the *current* study sample were exposed to more unconstrained than constrained instruction in both PreK and kindergarten (Weiland et al., under review). For literacy instruction specifically, children in our public PreK sample spent around 8.2 min per day in constrained literacy instruction, and 17 min in unconstrained language instruction skills. As a comparison, evidence from recent work done in Tennessee and North Carolina demonstrated the opposite trend, with 26.2 min spent in constrained instruction and 7.2 min in unconstrained instruction per day (Farran, Meador, Norvell, & Nesbitt, 2015). These findings suggest that the BPS PreK program may be unique in its instructional approach. Work examining links between enrollment in other PreK programs and constrained and unconstrained skills is needed to determine if this pattern of results extends to models that vary in their instructional foci and practices.

We must emphasize that our results in no way indicate that unconstrained skills are superior to constrained skills, or that PreK programs should focus all their attention on supporting unconstrained versus constrained skills in order to generate lasting benefits for students. In addition to these skills not being completely distinct from one another, a wide body of work has shown that skills that are considered more constrained are also foundational competencies that support the development of higher-order and more complex skills (Snow & Matthews, 2016). PreK programs must

continue to provide instruction to support these skills, which include key competencies like phonological awareness, letter and word identification, spelling, concepts about text and print, counting, and basic addition and subtraction. Findings do suggest, however, that programs should not restrict activities to constrained instruction and should instead expand their focus to strengthen *both* types of competencies. Considering additional supports to enhance unconstrained activities, like class discussions, supporting children to share and explain their thinking, comparing and contrasting, making predictions and inferences, developing concepts of measures, and summarizing and representing data (among others) may be warranted. Furthermore, evaluating which of the constrained skills that children have mastered can help teachers expose children to more challenging content, especially in the kindergarten year.

Links Between PreK and Constrained and Unconstrained Math Skills

There has also been limited research to date to suggest that one would be likely to observe more lasting effects of a public PreK program on unconstrained versus constrained *math* skills. Recent work by Mattera, Jacob, and Morris (2018) examining the effects of the Building Blocks math curriculum, relative to business as usual PreK in 69 centers and schools serving primarily low-income students, demonstrated that the program showed lasting effects on the REMA—the measure of unconstrained math skills assessed in our paper—through the spring of kindergarten. In that study, however, there were no lasting effects of the program on our constrained math outcome—the WJ Applied Problems. The BPS prekindergarten uses the Building Blocks math curriculum. Thus, although the comparison conditions across these studies are different, findings taken together do suggest that models like these may have more lasting effects on unconstrained skills, perhaps less likely to develop in the absence of the program (Bailey et al., 2017).

Comparison of Public PreK and Non-Public PreK Results

An important contribution of this study was our ability to disaggregate students who attended public PreK from students who attended non-public PreK programs and a third group of students who did not attend a formal PreK program as 4-year-olds. Prior studies of the BPS prekindergarten program (Weiland et al., 2019) and other scaled PreK models (e.g., Lipsey et al., 2018; Puma et al., 2010) have typically compared students who enrolled in the program of interest to all other students who did not access the program. The experiences of students in comparison conditions can be diverse in auspice and quality, making it difficult to understand *to what* the program of interest is being compared. By disaggregating conceptually similar groups in the current study we address this limitation from past work. Findings revealed that enrollment in public PreK and non-public PreK was associated with unconstrained language skills at the end of kindergarten. Effect sizes for both groups in the spring of kindergarten were fairly similar, although effects were larger for the public PreK program. This result provides some preliminary descriptive evidence that sustained associations between PreK and unconstrained language skills may not be limited to the BPS public program.

In contrast, when predicting math outcomes, we found more lasting associations between public PreK and constrained and unconstrained math skills than for non-public PreK. Indeed, our finding that there was no lasting effect of non-public PreK on either math outcome in the spring of kindergarten does suggest that the math experiences of children in the public and non-public settings were qualitatively different, given that the Building Blocks math curriculum was implemented in public PreK settings. Complementary work on a subset of the non-public settings suggests that students in those centers were less likely to be exposed to math instruction in general, and unconstrained math instruction in particular (Weiland et al., under review). It may be that the non-public settings adopted the BPS curriculum in ways that ensured better-than-business-as-usual instruction to support vocabulary skills. In contrast, the math skills these programs supported were equally likely to develop for children who enrolled in non-public settings and those who stayed at home. Findings may reflect a larger focus in non-public settings on instruction to enhance language skills. Descriptive research from Morris, Mattera, and Maier (2016) found that PreK classrooms located in community-based organizations had a bigger differential in time spent on language and literacy versus math than PreK classrooms in public school settings.

Variation in Effects of Public PreK on Students' Skills

We fit a number of models to test whether effects of public and non-public PreK on constrained and

unconstrained skills varied by students' family income, race/ethnicity, and Dual Language Learner status. Taking into account the number of statistical tests we conducted, results suggest little variation in these associations. The benefits of PreK programs on unconstrained skills through the spring of kindergarten were similar across levels of family income, white and non-white students, and monolingual English versus DLL children. This pattern stands in contrast with prior work suggesting that the benefits of PreK programs may be larger for low-income, DLL, and non-white—particularly Hispanic—students (Bloom & Weiland, 2015; Phillips, Gormley, & Anderson, 2016). Indeed, earlier research has generally found evidence that PreK may act as a compensatory factor for students from low-income, non-white, and DLL families (Phillips, Lipsey, et al., 2017; Yoshikawa et al., 2013).

Even so, there was evidence that both public and non-public PreK had larger associations with vocabulary in the fall of kindergarten for DLL students. PreK programs may help these students strengthen skills—particularly English language skills and vocabulary—that they would not otherwise have developed at home. Indeed, quasi-experimental analyses have shown that the impacts of Head Start were primarily concentrated in groups of students who would have stayed at home, in the absence of Head Start (Feller, Grindal, Miratrix, & Page, 2016). We also found that the associations between public and non-public PreK and constrained and unconstrained math skills in the fall of kindergarten were slightly larger for higher-income versus lower-income students. Some extant work suggests that more advantaged students may be more likely to benefit from PreK models like Building Blocks (Morris et al., 2016). A meta-analysis by Simonsmeier et al. (2018) has shown that less cognitively demanding interventions show bigger effects for students with lower baseline skills, and more cognitively demanding interventions demonstrate larger benefits for students with higher baseline skills. Our finding that the language and vocabulary skills supported through the PreK program may have provided a compensatory structure for DLLs aligns with their conclusion, as do our findings that advantaged students reap more benefits from the cognitively demanding math instruction delivered through the Building Blocks curriculum. Even so, more work examining these associations in larger samples with more statistical power is needed. As noted, we did not observe variation in any effects in the spring of kindergarten, suggesting that the evidence for the constrained versus unconstrained skills *hypothesis* did not differ across particular subgroups of interest.

Strengths and Limitations

The core strengths of this study are its rich measurement of student outcomes, inclusion of a diverse sample of students, focus on a school district with a high-quality prekindergarten program and a range of comparison PreK options for students, ability to examine effects over time, controlling for a robust set of covariates measured through a range of rich data sources, and use of numerous robustness checks. Even so, there are some limitations that are important to note. First, the analysis is non-experimental and results cannot be interpreted causally. Although we control for a range of covariates and explored a series of robustness checks—including a propensity score matching approach adjusting for indicators of the quality of the 4-year-old year home learning environment—to examine the sensitivity of our findings to various specifications, we are unable to exclude the possibility of omitted variables that could confound our findings. Future experimental work that is able to generate similar groups prior to exposure to PreK (with respect to both demographics and exposure to variation in the home learning environment) is needed to rigorously estimate effects of PreK on constrained and unconstrained skills across time. In addition, we examined associations between enrollment in BPS public PreK and students' outcomes in the fall of kindergarten and were unable to more precisely estimate the effect of public and non-public PreK outcomes assessed in the spring of the PreK year. Second, although study participants were randomly sampled from the broader BPS population, the results from the current study may not be wholly generalizable to this group. Furthermore, aside from the students enrolled in the CBOs receiving supports from the school district, we lack rich information on the students in the comparison conditions and the supports they received during their 4-year-old year. Future work that documents the experiences of students across PreK attender groups could help fill these gaps.

Fourth, the group of students who did not attend any formal PreK program during their 4-year-old year split is fairly small relative to the groups of public and non-public PreK attenders. The small sample size posed a challenge to our models examining variation in associations by family income, race/ethnicity, and DLL status. For example, we were unable to examine variation in associations for

different subgroups of non-White students. Future work with more power to examine subgroup effects is needed. Fifth, our measure of unconstrained math skills—the REMA *t*-score—does include some assessed items that could be conceptualized as constrained math skills and the correlations between measures of constrained and unconstrained math skills were larger than the correlations between constrained and unconstrained language and literacy skills. The field as a whole currently has less conceptual clarity about distinctions in constrained and unconstrained skills in the domain of the math. Related, the bulk of students in the public PreK program (and some students in the non-public PreK program) did experience the Building Blocks curriculum during the PreK year which may be particularly aligned with the REMA outcome, as they were developed by the same authors. Future measurement work in this area will help inform ongoing research to examine constrained versus unconstrained math skills in studies of early childhood and K-12 education with a core goal being to reduce the intercorrelation in the assessments of constrained and unconstrained math skills.

Next, it is impossible for us to disentangle whether the lasting associations are fully attributed to the public PreK program or to the effort to align instruction across public PreK and kindergarten. Although we were able to answer our key question and show that effects were more likely sustained for unconstrained versus constrained skills, further experimental work disaggregating the effects of PreK and alignment is warranted. In this study, we were also only able to examine our hypotheses through the spring of kindergarten. As other studies in the field have found, it is possible that patterns of convergence could emerge in later elementary school grades. Future research examining later time points—particularly third grade—will provide needed information on the extent to which associations between enrollment in public PreK and skills are sustained across time. Finally, we recognize that the BPS prekindergarten program is somewhat unique in the field, given its focus on supporting unconstrained skills and its link to a kindergarten program that is intentionally aligned. As such, our findings may not generalize across all PreK programs. Further research involving a range of scaled PreK models is thus warranted.

Implications for Policy and Practice

Despite the descriptive nature of our study findings, results generate important implications for future research, practice, and policy. First, we provide evidence showing that the effects of PreK on unconstrained skills may be more likely to be sustained than the effects of PreK on constrained skills. This underscores a need for future evaluations of PreK programs to continue to explore *both* constrained and unconstrained skills across time. Second, our results point to the need for more robust measures of unconstrained language and literacy and math skills as children move out of PreK and into elementary school. For example, there is a component of the BPS prekindergarten curriculum that involves children writing and telling stories to their peers, and then actually acting them out for their class. The documentation collected as part of this component could be reviewed and coded to assess unconstrained skills like complexity of vocabulary and critical thinking. Further integration of assessments of constrained and unconstrained skills into PreK impact studies would be a valuable contribution to the field.

If a key goal of PreK models is to have lasting benefits for student participants, our findings suggest the importance of implementing rich curricula focused on language and literacy and math that include instruction supporting both constrained and unconstrained skills. This type of program model may stand in opposition to the dominant narrative around preschool curricula as either academic and focused on constrained skills *or* play-based and focused on unconstrained skills (Weiland, 2018). As the Boston example illustrates, instruction targeting both unconstrained and constrained skills can co-occur and likely should, given that both types of skills are foundational. Policymakers and practitioners interested in implementing PreK models that promise lasting benefits for students may consider such program models to achieve these goals.

References

Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, *145*, 1103. https://doi.org/10.1037/bul0000212

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*, 7–39. https://doi.org/10.1080/19345747.2016.1232459

Barrueco, S., Lopez, M., Ong, C., & Lozano, P. (2012). *Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures*. Baltimore, MD: Brookes.

Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *The Reading Teacher*, *55*, 10–20. Retrieved from https://www.jstor.org/stable/20205005

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal*, *111*, 7–34. https://doi.org/10.1086/653468

Bloom, H. S., & 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. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2594430

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *29*, 3–20. https://doi.org/10.2307/749715

Chapin, S., & O'Connor, C. (2012). Project Challenge: Using challenging curriculum and mathematical discourse to help all students learn. In *High expectation curricula: Helping all students succeed with powerful learning* (pp. 113–127). https://doi.org/10.3102/978-0-935302-43-1_9

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to kindergarten: A new plan to combat inequality*. New York, NY: Russell Sage Foundation. https://doi.org/10.1086/697591

Choi, J. Y., Elicker, J., Christ, S. L., & Dobbs-Oates, J. (2016). Predicting growth trajectories in early academic learning: Evidence from growth curve modeling with Head Start children. *Early Childhood Research Quarterly*, *36*, 244–258. https://doi.org/10.1016/j.ecresq.2015.12.017

Choi, J. Y., Jeon, S., & Lippard, C. (2018). Dual language learning, inhibitory control, and math achievement in Head Start and kindergarten. *Early Childhood Research Quarterly*, *42*, 66–78. https://doi.org/10.1016/j.ecresq.2017.09.001

Claessens, A., Engel, M., & Curran, F. C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, *51*, 403–434. https://doi.org/10.3102/0002831213513634

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*, 136–163.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Math Assessment. *Educational Psychology Review*, *28*, 457–482. https://doi.org/10.1080/01443410701777272

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge. https://doi.org/10.4324/9780203774441

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., … Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, *38*, 85–99. https://doi.org/10.3102/0013189x09332373

Duncan, S. E., & DeAvila, E. A. (1998). *PreLAS*. Monterey, CA: CBT McGraw Hill.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Pearson Assessments. https://doi.org/10.1037/t15144-000

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M., Crawford, G. M., … Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *25*, 177–193. https://doi.org/10.1016/j.ecresq.2009.10.003

Engel, M., Claessens, A., Watts, T., & Farkas, G. (2016). Mathematics content coverage and student learning in kindergarten. *Educational Researcher*, *45*, 293–300. https://doi.org/10.3102/0013189x16656841

Farran, D. C., Wilson, S. J., Meador, D., Norvell, J., & Nesbitt, K. (2015). *Experimental evaluation of the tools of the mind Pre-K curriculum*. Technical report. Working paper, Peabody Research Institute.

Feller, A., Grindal, T., Miratrix, L., & Page, L. C. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care type. *The Annals of Applied Statistics*, *10*, 1245–1285. https://doi.org/10.1214/16-aoas910

Good, R. H., Kaminski, R. A., Cummings, K. D., Dufour-Martel, C., Petersen, K., Powell-Smith, K., … Wallin, J. (2011). *DIBELS next assessment manual*. Eugene, OR: Dynamic Measurement Group.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*, 872. https://doi.org/10.1037/0012-1649.41.6.872

Guerrero-Rosada, P., Weiland, C., McCormick, M. P., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2020). Process quality and children's language, math, and executive function gains in prekindergarten: A replication and extension study. *Early Childhood Research Quarterly*, *54*, 1–12. https://doi.org/10.1016/j.ecresq.2020.07.009

Guo, S., & Fraser, M. W. (2015). Propensity score matching and related models. *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781071802854.n5

Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, *32*, 60–79. https://doi.org/10.1016/j.ecresq.2014.12.005

Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010). *The New Mexico PreK evaluation: Impacts from the fourth year (2008–2009) of New Mexico's state-funded PreK*

program. New Brunswick, NJ: National Institute for Early Education Research. https://doi.org/10.1037/e584752012-278

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. London, UK: Cambridge University Press.

Li, W., Leak, J., Duncan, G. J., Magnuson, K., Schindler, H., & Yoshikawa, H. (2016). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program*. UCI SoE working paper, Graduate School of Education, University of California, Irvine, CA.

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, *45*, 155–176. https://doi.org/10.1016/j.ecresq.2018.03.005

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*, 159–208. https://doi.org/10.1162/qjec.122.1.159

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, NY: MDRC.

McCormick, M., Hsueh, J. A., Weiland, C., & Bangser, M. (2017). *The challenge of sustaining preschool impacts: Introducing ExCEL P-3, a study from the Expanding Children's Early Learning Network*. New York, NY: MDRC.

McCormick, M. P., Weiland, C., Hsueh, J., Maier, M., Hagos, R., Snow, C., … Schick, L. (2020). Promoting content-enriched alignment across the early grades: A study of policies & practices in the Boston Public Schools. *Early Childhood Research Quarterly*, *52*, 57–73. https://doi.org/10.1016/j.ecresq.2019.06.012

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.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*, 184–202. https://doi.org/10.1598/rrq.40.2.3

Phillips, D., Gormley, W., & Anderson, S. (2016). The effects of Tulsa's CAP Head Start program on middle-school academic outcomes and progress. *Developmental Psychology*, *52*, 1247–1261. https://doi.org/10.1037/dev0000151

Phillips, D., Johnson, A., Weiland, C., & Hutchison, J. E. (2017). *Public preschool in a more diverse America: Implications for next-generation evaluation research*. Ann Arbor, MI: Poverty Solutions.

Phillips, D., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., & Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement*. Issues in Pre-Kindergarten Programs and Policy (pp. 19–30).

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Baltimore, MD: Paul H Brookes. https://doi.org/10.1037/t08945-000

Powell, D. R., Son, S. H., File, N., & San Juan, R. R. (2010). Parent–school relationships and children's academic and social outcomes in public school pre-kindergarten. *Journal of School Psychology*, *48*, 269–292. https://doi.org/10.1016/j.jsp.2010.03.002

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of US fourth- and eighth-grade students in an international context*. NCES 2013–009. Jessup, MD: National Center for Education Statistics.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., … Spier, E. (2010). *Head Start impact study*. Final report. Executive summary. Administration for Children & Families. Retrieved from http://files.eric.ed.gov/fulltext/ED507847.pdf

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, *2*, 2332858416657343. https://doi.org/10.1177/2332858416657343

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*, 958. https://doi.org/10.1037/a0015861

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. *Oxford Handbook of Numerical Cognition*, 1118–1134. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bi-directional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*, 587–597. https://doi.org/10.1007/s10648-015-9302-x

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge. https://doi.org/10.4324/9780203883785

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *5*, 105–135. https://doi.org/10.1080/19345747.2011.627980

Schickedanz, J. A., & Dickinson, D. K. (2004). *Opening the world of learning: A comprehensive early literacy program*. Parsippany, NJ: Pearson Early Learning.

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations between conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*, 1525–1538. https://doi.org/10.1037/a0024997

Schrank, F. A., McGrew, K. S., Ruef, M. L., & Alvarado, C. G. (2005). Batería III Woodcock-Muñoz™. *Assessment Service Bulletin*. https://doi.org/10.1037/e537522009-008

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin. https://doi.org/10.1086/345281

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2018). *Domain-specific prior knowledge and learning: a meta-analysis*. Research synthesis 2018, Trier, Germany. https://doi.org/10.23668/psycharchives.844

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades reading and language in the early grades. *Future of Children*, *26*, 57–74. https://doi.org/10.1353/foc.2016.0012

Weiland, C. (2018). Commentary: Pivoting to the "how": Moving preschool policy, practice, and research forward. *Early Childhood Research Quarterly*, *45*, 188–192. https://doi.org/10.1016/j.ecresq.2018.02.017

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*, 2332858418757735. https://doi.org/10.1177/2332858418757735

Weiland, C., Moffett, L., Guerrero-Rosada, P., Weissman, A., Zhang, K., Maier, M., … Sachs, J. (under review). Is child-level measurement the key to improving the predictive validity of observational measures in early education classrooms?. *Journal of Research on Educational Effectiveness*.

Weiland, C., Unterman, R., & Shapiro, A. (in press). The kindergarten hotspot: Literacy skill convergence between Boston Prekindergarten enrollees and non-enrollees. *Child Development*.

Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2019). The effects of enrolling in oversubscribed prekindergarten programs through third grade. *Child Development*. https://doi.org/10.1111/cdev.13308

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

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. https://doi.org/10.1002/pam.20310

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside. https://doi.org/10.1007/springerreference_180723

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Policy Brief: Society for Research in Child Development and Foundation for Child Development. Retrieved from http://disde.minedu.gob.pe/handle/123456789/4015

Yudron, M., Weiland, C., & Sachs, J. (2016). *BPS K1DS: Piloting the Boston Public Schools' prekindergarten model in community-based organizations*. Boston, MA: Boston Public Schools.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Supplementary information on data, methodology, results, and robustness checks