Does Kindergarten Instruction Matter for Sustaining the Prekindergarten (PreK) Boost? Evidence From Individual- and Classroom-Level Survey and Observational Data

Meghan P. McCormick[1], Mirjana Pralica[1], Christina Weiland[2], JoAnn Hsueh[1], Lillie Moffett[2], Paola Guerrero-Rosada[2], Amanda Weissman[2], Kehui Zhang[2], Michelle F. Maier[1], Catherine E. Snow[3], Emily Davies[1], Anne Taylor[4], and Jason Sachs[4]

[1] MDRC, Families' Well-Being and Children's Development Policy Area, New York, New York, United States; [2] School of Education, University of Michigan; [3] Graduate School of Education, Harvard University; [4] Department of Early Childhood, Boston Public Schools

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. Work done by Amanda 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 Boston Public Schools (BPS) Department of Early Childhood coaches and staff, the BPS Department of Research, the MDRC team, the Harvard Graduate School of Education team, and the University of Michigan team. Thank you to members of the Consortium on Early Childhood Intervention Impact who provided helpful feedback and comments on a draft of this article. Special thanks to the students and teachers who participated directly in the research. This study was not preregistered.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

2022, Vol. 58, No. 7, 1298–1317
https://doi.org/10.1037/dev0001358

Does Kindergarten Instruction Matter for Sustaining the Prekindergarten (PreK) Boost? Evidence From Individual- and Classroom-Level Survey and Observational Data

Meghan P. McCormick[1], Mirjana Pralica[1], Christina Weiland[2], JoAnn Hsueh[1], Lillie Moffett[2], Paola Guerrero-Rosada[2], Amanda Weissman[2], Kehui Zhang[2], Michelle F. Maier[1], Catherine E. Snow[3], Emily Davies[1], Anne Taylor[4], and Jason Sachs[4]

[1] MDRC, Families' Well-Being and Children's Development Policy Area, New York, New York, United States
[2] School of Education, University of Michigan
[3] Graduate School of Education, Harvard University
[4] Department of Early Childhood, Boston Public Schools

The sustaining environments hypothesis theorizes that the lasting effects of PreK programs are contingent on the quality of the subsequent learning environment in early elementary school. The current study tests this theory by leveraging data from students (N = 462) who did and did not enroll in the Boston Public Schools (BPS) prekindergarten (PreK) program as well as features of their kindergarten instruction measured at the child- and classroom-levels using surveys and observations. Taken together, findings revealed limited evidence for the sustaining environments hypothesis. The bulk of the results were null, indicating that in general, associations between enrollment in BPS PreK and language, literacy, and math skills through the spring of kindergarten did not vary by kindergarten instructional experiences. When examining distinct types of instructional experiences, there were some inklings that *child-level* observational measures of kindergarten learning experiences—particularly those capturing constrained versus unconstrained instruction—were more predictive of PreK persistence than observed global classroom quality measures or survey-based measures of advanced instruction. However, these associations were not always specific to outcomes matching the content delivered during this instruction (math vs. literacy), consistent with the possibility of either cross-domain effects or that instructional variables are proxies for more general instructional practices. Findings for future research and theory are discussed.

Keywords: academic skills, convergence, fadeout, PreK

Supplemental materials: https://doi.org/10.1037/dev0001358.supp

Across rigorous studies of early care and education programs conducted decades ago and in recent years, PreK participants outperform nonparticipants on assessments of language, literacy, and math skills prior to the start of kindergarten. However, the PreK boost diminishes in elementary school, sometimes partially and sometimes fully (Yoshikawa et al., 2016). This phenomenon has been described as both "fadeout"—where the benefits of PreK lessen over time—and "convergence"—where children who do not attend PreK eventually catch up to their peers (McCormick et al., 2017). Consistent evidence of this trend has ignited policy debates questioning the utility of public investments in high-quality PreK. However, with federal policy priorities increasingly shifting toward expanding universal PreK for all children in the United States, there is a pressing need to better understand the conditions under which the effects of these investments in early childhood education are most likely to persist across time (Stipek et al., 2017).

This article was published Online First May 5, 2022.
Mirjana Pralica https://orcid.org/0000-0002-8676-9691
Lillie Moffett https://orcid.org/0000-0003-1233-5185
Paola Guerrero-Rosada https://orcid.org/0000-0002-0310-5890
Amanda Weissman https://orcid.org/0000-0001-5381-4516
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. 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 Boston Public Schools (BPS) Department of Early Childhood coaches and staff, the BPS Department of Research, the MDRC team, the Harvard Graduate School of Education team, and the University of Michigan team. Thank you to members of the Consortium on Early Childhood Intervention Impact who provided helpful feedback and comments on a draft of this article. Special thanks to the students and teachers who participated directly in the research. This study was not preregistered. Data are not yet publicly available but will be shared as a restricted access file—available via application—after the conclusion of the grant period in 2023.

Correspondence concerning this article should be addressed to Meghan P. McCormick, MDRC, Families' Well-Being and Children's Development Policy Area, 200 Vesey Street 23rd Floor, New York, NY 10281, United States. Email: Meghan.mccormick@mdrc.org

The "sustaining environments" hypothesis has emerged as a leading theory to explain the pattern of PreK fadeout/convergence (Abenavoli, 2019). As described by Bailey, Duncan, et al. (2020), this conjecture refers to the "popular idea, stemming from theories in developmental, cognitive, and educational psychology, that the long-term success of early educational interventions is contingent on the quality of the subsequent learning environment" (p. 2). Empirical evidence on this theory is mixed (e.g., Bailey, Jenkins, et al., 2020), perhaps because elementary school experiences tend to be measured globally using metrics like observed quality of teacher-child interactions (Carr et al., 2019), teacher reports of instructional content (Jenkins, Watts, et al., 2018; Magnuson et al., 2007), or standardized test scores at the school level (Pearman et al., 2020; Unterman & Weiland, 2020). These global measures may not account for individual-level differences in the learning experiences of children within the same classrooms (Connor et al., 2009). In addition, global classroom measures that focus on assessing the quality of *interactions* between teachers and students may not fully capture key indicators of instructional quality—such as time spent in direct instruction, activity settings, and the type and level of instruction—hypothesized to drive gains in children's academic development (Burchinal, 2018).

The current study aims to build on the existing literature exploring the sustaining environments hypothesis by leveraging information on children's instructional experiences in kindergarten, measured both globally at the classroom-level and individually at the student-level. Global measures capture observed process and instructional quality as well as the extent to which instructional content was advanced or more basic. Student-level observations assess the amount of time individual children were exposed to different activity settings (e.g., centers, small groups, whole group, individual time); time spent in active instruction and noninstruction; and the amount of time spent in different content domains. These data are particularly unique because, within this last grouping, they describe instruction supporting children's *constrained* and *unconstrained* language/literacy and math skills. Rooted in theory developed by Paris (2005), instruction focused on unconstrained skills enhances broadband competencies like vocabulary, problem solving, and critical thinking that are acquired gradually and can be difficult to measure. More constrained instruction supports skills like letter naming and counting skills (among others) that are more finite and directly teachable. Although both are critical for children's success, emerging evidence suggests that the benefits of early childhood educational interventions may be more likely to persist when instruction focuses explicitly on unconstrained skills to some degree (McCormick et al., 2021).

We conduct this study using data on students who did and did not enroll in the Boston Public Schools prekindergarten (BPS PreK) program during the 2016–2017 academic year.[1] Importantly, an earlier study (focused on children who enrolled in the program in 2009–2010) using an age-based regression discontinuity design found sizable casual impacts of the BPS PreK program on children's language, literacy, and math skills at the start of kindergarten, indicating evidence of a short-term "boost" to be sustained (Weiland & Yoshikawa, 2013). The current study uses a descriptive design to build on that earlier work and study the program during a time when the majority of public schools in BPS implemented curricula and professional development to align instruction across PreK and kindergarten McCormick et al. (2020). By using both student- and classroom-level data from kindergarten to assess instructional experiences in this new cohort of BPS students, our findings stand to both build on the prior regression discontinuity study and produce new knowledge on the utility of student-level observations to explain the fadeout/convergence puzzle.

PreK Fadeout/Convergence and the Sustaining Environments Hypothesis

The short-term benefits of PreK programs attended during the four-year-old year universally become much smaller as children move through elementary school (Lipsey et al., 2018; Puma et al., 2010; Wong et al., 2008; Yoshikawa et al., 2013). A recent meta-analysis found that about half of this eventual convergence on cognitive outcomes occurs during kindergarten and then by about half again by the end of second grade (Li et al., 2020). Bioecological theory (Bronfenbrenner & Morris, 1998) suggests that children's microsystems, such as their PreK and kindergarten environments, not only have direct influences on their development, but also interact with one another to further affect children's outcomes. Even if children experience a short-term boost in academic skills as the result of a high-quality PreK program, subsequent low-quality environments may undermine those gains.

For example, children can encounter a situation labeled by Bailey and colleagues (2016) as "constraining content" wherein elementary school settings repeat content learned in PreK. Research by Engel et al. (2013) using nationally representative data has shown that the majority of kindergarten students are exposed to content they already know. In contrast, when children who attend PreK transition into high-quality kindergarten settings—hypothesized to include such features as highly effective teaching (e.g., Pearman et al., 2020), exposure to advanced instructional content (Jenkins, Duncan, et al., 2018; Bassok et al., 2019), high-quality interactions between teachers and students (e.g., Carr et al., 2019), and high levels of overall school quality (Unterman & Weiland, 2020)—ecological theory would suggest that those contexts present positive independent and interactive influences that further stand to sustain or magnify the benefits of PreK across time.

Yet, the empirical evidence for the sustaining environments theory is actually quite mixed, with studies reporting results that both align with and contradict this hypothesis. For example, Swain et al. (2015) analyzed follow-up data from the Tennessee Voluntary PreK study and found that students who were assigned to the PreK program and then enrolled in a first grade classroom with a highly effective teacher—measured on a state-developed global teacher rating system—demonstrated sustained benefits of PreK on a composite measure of early math and literacy skills. In addition, Unterman and Weiland (2020) leveraged data from a lottery-based study to estimate variation in impacts of the BPS PreK program on third grade standardized test scores. They found quasi-experimental evidence that the impacts of the program were sustained—with large effect sizes (ELA, $d = .45$; math, $d = .66$)—for compliers who were randomly assigned to and enrolled in a high-

[1] The preschool program in the Boston Public Schools describes itself as BPS prekindergarten. For the sake of brevity, we describe the program in this article as BPS PreK.

quality school compared with compliers assigned to the control group who did not access the program.

In contrast, other studies have found that the benefits of PreK tend to fadeout/converge regardless of the quality of the subsequent learning environment. As argued by Jenkins, Watts, and colleagues (2018), although higher-quality instruction subsequent to PreK may improve the academic development of PreK attenders, this instruction may boost the learning of children who did not attend PreK *even more*. Work using nationally representative data has shown that all children benefit from exposure to advanced content in the early grades regardless of whether they attended PreK or not, and do not benefit from basic content coverage (Claessens et al., 2014). Even if PreK attenders continue to build on their gains when exposed to high-quality instruction in elementary school, the convergence will still occur if those high-quality environments have an even larger positive influence on children who did not attend PreK. Work leveraging data from trials of Head Start (Puma et al., 2010) and the Building Blocks PreK math curriculum (Clements et al., 2008) found evidence for this theory, reporting that the quality of kindergarten and first grade instruction—assessed using both teacher reports of the frequency of language/literacy practices and an observational measure of math quality and dosage—did not sustain the initial effects of the early intervention (Jenkins, Watts, et al., 2018).

Still another body of literature suggests that the benefits of PreK are actually *more* likely to converge when elementary school settings are of higher quality. For example, a descriptive analysis using nationally representative data (Magnuson et al., 2007) found that PreK convergence was *most likely* for students who experienced elementary school classrooms that spent *more* time on reading instruction, measured with one teacher-reported survey measure. In contrast, the benefits of PreK persisted when students enrolled in elementary school settings that spent *less* time on reading instruction and had larger class sizes. Classrooms that spent more time on reading helped non-PreK attenders catch up to their peers. Follow-up data from the randomized trial of Head Start REDI found that program impacts on most social-emotional outcomes were more likely to be sustained when children were in kindergarten classrooms with a higher proportion of students scoring lower on standardized assessments of academic skills (Bierman et al., 2014). However, benefits of Head Start REDI on one outcome —competent problem solving—were more likely to persist for children in kindergarten classrooms with higher-quality teacher-child interactions. Even so, findings taken together suggest that sustaining environments—when specifically defined as contexts that support the PreK advantage—may actually be characterized by contexts that are *less* supportive of both the counterfactual group and the group of students who did attend PreK.

A recent study by Bailey and colleagues (2016) has further raised the possibility that general quality *on its own* is not necessarily a key support needed to sustain the benefits of early interventions. In their study, the authors leveraged data from a randomized control trial of the Building Blocks PreK math curriculum and found that 72% of the fadeout/convergence effect was attributed to preexisting differences between children in the treatment and control groups with the same level of achievement at posttest. Results from a recent article by Ansari et al. (2020) examining a scaled PreK program in a large suburban school district align with these findings and the importance of preexisting student characteristics in observing patterns of persistence. These studies suggest the need for more *tailored curricular interventions*—rather than general quality supports—to sustain the PreK boost. For example, Mattera et al. (2018) found that the impacts of Building Blocks were more likely to be sustained when students were exposed to tailored kindergarten intervention—clubs where students did hands-on math activities—that explicitly aligned with the instructional content of PreK programming.

Measuring Sustaining Environments

There are a number of *specific* instructional activities within elementary school classroom microsystems that have been linked to students' academic outcomes (Burchinal, 2018; Weiland et al., 2018) and shown to be malleable to classroom-level intervention through implementation of particular curricula, coaching, and training (Weiland et al., 2018). A recent review by Burchinal (2018) argues that the content and activity settings (e.g., small groups that support improved differentiation) that students are exposed to in early childhood may be important predictors of academic skills. However, prior work has largely used global measures (with one score per classroom) to assess features, such as instructional content (e.g., amount of time spent in content domains, instruction vs. non-instruction; Farran & Bilbrey, 2004), the level for example, (teacher- or child-managed) and type of instruction (Connor et al., 2004), and activity settings (time spent in whole group, small group, centers, individual time; Fuligni et al., 2012).

Type of instruction can further be described as supporting children's more *constrained* or *unconstrained* skills. Following constrained skill theory (McCormick et al., 2021; Paris, 2005; Snow & Matthews, 2016), early language, literacy, and math skills can be thought of as existing on continuum of constrained to unconstrained. More *constrained skills* are those that can be directly taught and assessed and are typically mastered by most children in early elementary school. Instructional activities supporting more constrained literacy skills would include time spent on phonological awareness, basic print and text concepts, and reading fluency, whereas more constrained math activities would include number sense and concepts and basic operations. As argued by Snow and Matthews (2016), more *unconstrained skills* are those that develop more gradually over time but are hypothesized to benefit students to a greater extent in the long run. Instructional activities supporting unconstrained language skills would include time spent exposing children to new vocabulary and developing reading comprehension competencies, whereas unconstrained math activities would include data analysis and measurement (see Table S2 in the online supplemental materials for a full set of examples of constrained and unconstrained language/literacy and math skills). As argued by Paris (2005) and Snow and Matthews (2016), it is likely important for PreK curricula to balance instruction with similar amounts of focus paid to both types of skills to enhance the full range of competencies critical for continued learning and development. Most assessments of children's skills used in PreK and kindergarten capture more constrained skills and teachers may be motivated to spend more time teaching those competencies because they will be held accountable for them when child assessment data are reviewed (McCormick et al., 2021). Yet, to our knowledge there is no research to date that has explicitly measured individual children's exposure to learning activities conceptualized

as being constrained or more unconstrained. Further work measuring and describing more constrained and more unconstrained instruction as children move through the early grades is needed.

Importantly, when captured globally, at the classroom or school level, any measure of PreK quality—including the general pattern of interactions between students and teachers as well as the amount of time that students experience different types of instruction and spend in different activities settings (e.g., small group, whole group, centers)—may also mask *within-classroom variation* in students' learning experiences. This is a potential limitation of existing measures given work by Connor and colleagues (2020) finding that students' learning opportunities (e.g., play, literacy, math) and the frequency of specific discourse moves children and teachers use (e.g., child-directed vs. teacher-director) vary significantly for individual students within the same classroom. These empirical findings align with Vygotsky's theory on the zone of proximal development (Vygotsky, 1978), suggesting that learning experiences might differ across children in the same classroom—depending in part on whether they attended PreK or not—because the teacher is responding to individual children's developmental levels and skills when they enter kindergarten. Children who attend PreK and have already developed foundational literacy skills—such as letter and word knowledge—may engage in a more advanced set of literacy activities in kindergarten than a child who did not attend PreK and has yet to master those competencies. Global measures of subsequent classroom- and school-quality, however, are unable to capture those differences in students' within-classroom learning opportunities and may mask variation in how key instructional aspects of elementary school learning environments (e.g., time spent in content domains, type of instruction, time in activity settings) do or do not sustain the PreK boost. It may be fruitful to consider both global quality measures and individual-level experiences—such as time in small groups that better support scaffolding and differentiation—when exploring the sustaining environments hypothesis.

The Current Study

The current study examines the extent to which associations between enrollment in the BPS PreK program (McCormick et al., 2020) and children's language/literacy and math skills at the end of kindergarten do or do not vary by students' instructional experiences in kindergarten, measured using individual observations, global observations of classroom process and instructional quality, and teacher surveys of classroom practices. An earlier study examining this sample found that, compared with no center-based PreK, there were positive associations between enrollment in BPS PreK and children's language (d = .57), literacy (d = .55–.63), and math skills (d = .40–.65) in the fall of kindergarten. In the spring of kindergarten there was complete convergence on measured literacy skills, a halved effect size for the more constrained measure of math skills (reduced from .65 in the fall to .31 in the spring), and greater evidence of persistence for language skills (d = .48) and the more unconstrained measure of math skills (d = .30 in spring compared with .40 in fall). The current study builds on this work and answers the following research questions:

1. To what extent is students' *individual time* spent in instruction and non-instruction, different activity settings, and different types of language/literacy and math instruction associated with PreK persistence in students' language, literacy, and math skills in the spring of kindergarten?

2. To what extent are the *global quality* of teacher-child interactions in kindergarten (emotional support, classroom organization, instructional support) associated with PreK persistence in students' language, literacy, and math skills in the spring of kindergarten?

3. To what extent is the *level of instructional content* in kindergarten classrooms (basic versus advanced) associated with PreK persistence in students' language, literacy, and math skills in the spring of kindergarten?

Although our study is exploratory, results enhance understanding of the sustaining environments hypothesis and help to better understand the PreK fadeout/convergence puzzle.

Method

Participants

The sample for the current study consists of 462 students attending the BPS kindergarten program during the 2017–2018 year who were enrolled in a classroom that participated in data collection activities. These students are a subsample of the broader population of those recruited into the study across the 2016–2017 (PreK; N = 388) and 2017–2018 (kindergarten; N = 176 additional students) academic years. In the fall of 2016, we recruited students from 41 public BPS PreK classrooms and 10 community-based (CBO) classrooms, nested within 20 public schools and 10 CBO centers. We then recruited additional students in the fall of 2017 who had not attended BPS PreK in the public schools from 80 kindergarten classrooms in 39 public schools. Of the full student sample, there are N = 284 students who attended the BPS PreK program and enrolled in the study in 2016–2017. Students who were added to the sample in kindergarten (N = 178) experienced a range of settings in the year prior to kindergarten–38% of them stayed home with a parent, family member or other adult, or attended a home-based daycare and the remaining 62% enrolled in a nonpublic PreK program for at least some of the time. A subsample of study participants (N = 21 schools, N = 49 classrooms) participated in videotaping activities that yielded data for our individual-level measures of children's learning activities (N = 385 students) and observed classroom quality (N = 402 students). Of the student sample, 431 were enrolled in a kindergarten classroom where the teacher completed a survey reporting on their instructional practices.

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 and the students were diverse with respect to racial/ethnic background and parental education, among other characteristics. As described more fully in McCormick et al., 2021, the public schools in the sample are representative of the broader population of BPS elementary schools and the students in our analytic sample differed slightly from the broader sample of kindergarten students in BPS.

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

Characteristic	BPS PreK		Other care			
	M or %	*SD*	*M* or %	*SD*	*M* Diff.	Stat. Sig.
Student demographic characteristics						
Child age	5.48	0.28	5.44	0.31	0.04	
Eligible free/red. lunch	0.59	0.49	0.73	0.45	−0.14	*
Female	0.48	0.50	0.54	0.50	−0.06	
Dual Language Learner	0.55	0.50	0.58	0.49	−0.03	
Asian	0.18	0.38	0.23	0.42	−0.05	
Black	0.18	0.39	0.28	0.45	−0.10	*
Hispanic	0.32	0.47	0.33	0.47	−0.01	
Other race	0.06	0.23	0.01	0.11	0.05	*
White	0.26	0.44	0.15	0.35	0.11	**
Parent demographic characteristics						
Age of mother when child was first born	27.20	6.94	24.67	6.12	2.53	***
Total household size	4.28	1.23	4.34	1.46	−0.06	
Parent works full-time	0.93	0.26	0.82	0.38	0.11	***
Parent married	0.59	0.49	0.46	0.50	0.13	*
Parent age	37.19	7.22	35.58	7.01	1.61	*
Parent education (categorical variable)[a]						
High school diploma/GED or less	0.29	0.45	0.39	0.49	−0.10	**
Some college/technical or vocational cert.	0.26	0.44	0.36	0.48	−0.10	
4-year degree	0.20	0.40	0.12	0.33	0.08	
>4-year degree	0.25	0.43	0.14	0.34	0.11	

Note. Diff. = difference; Stat. sig. = Statistical significance; Eligible free/red. lunch = eligible for free or reduced price lunch; BPS = Boston Public Schools. BPS PreK, $N = 284$, Other care, $N = 178$. Independent samples t tests used to test for differences between binary variables.
[a] Chi-square used to test for differences between parental education groups.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Procedure

The Institutional Review Boards at the lead organization for this study approved the human subjects plan prior to the commencement of study activities. The project name is ExCEL P-3: Promoting Sustained Gains from Preschool to Third grade and the study was approved by the MDRC IRB (approval number 860661-2).

School and Classroom Recruitment

In 2016; before the start of the PreK, we randomly selected 25 public schools to participate in the study from the 76 schools in the district offering the public school BPS PreK 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. All PreK classrooms in the current study used the BPS *Focus on K1* curriculum, a synthesis of two evidence-based curricula: an adapted version of Opening the World of Learning (Schickendanz & Dickinson, 2005), a language and literacy curriculum that includes a social-emotional skills component in each unit, and Building Blocks, an early mathematics curriculum that also promotes language development by requiring children to explain their mathematical reasoning verbally (McCormick et al., 2017). Teachers received training and coaching as needed depending on their tenure with their program and their school's NAEYC accreditation status. The BPS PreK program is implemented in public schools and colocated with elementary grades, a structural feature theorized to support program quality (Choi et al., 2016).

We randomly selected 10 of the 11 CBOs in Boston implementing the BPS PreK model to participate in the study and they all agreed. These CBOs were connected to the BPS Department of Early Childhood and were receiving training and coaching to implement the BPS curricular model during the 2016–2017 year. This was a first step in Boston's move to a universal mixed-delivery PreK system—with the BPS curriculum and professional development model offered in both public schools and CBOs—available to all eligible 4-year-olds in the city. We initially considered treating students enrolled in these CBOs as having experienced the BPS PreK program, given similarities in curricula and teacher professional development across them. However, there were a number of structural and qualitative differences between the public school BPS PreK program described above and the program being implemented in CBOs that forced us to reconsider that approach. First, CBOs implementing the BPS PreK curricula continued to operate outside the purview of the broader public school system and were not colocated with elementary schools. Children who attended CBO programs shifted to a different school location for kindergarten, while children in the public school BPS PreK program experienced (for the most part) colocation across years. Further, the teachers in CBO centers were not subject to the same educational requirements (masters degree within 5 years) as public school teachers, nor where they paid on the same scale. In addition, descriptive statistics revealed that the gains in academic skills that CBO students made during the PreK year were more similar in magnitude to those made by students who attended center-based PreK outside of the BPS public school PreK program. As such, we decided not to include students who attended the CBO programs implementing the BPS PreK curricula as part of the group who did enroll in the public school BPS PreK program. Instead, they are part of the comparison group of students who experienced any

care situation other than the BPS PreK program during the 2016–2017 school year.

We asked all PreK 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 and 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. Ninety-five percent of kindergarten teachers agreed to participate in the study in general (e.g., doing surveys and allowing their students to be assessed). However, given resource constraints coupled with some teachers not agreeing to videotaping in their classrooms, we were only able to collect videotaped data from 49 classrooms in kindergarten and collected surveys from $N = 80$ teachers (as described above).

Student Recruitment

We attempted to collect active consent for all PreK 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, we randomly selected 50% (~6–10 per classroom) to participate in student-level data collection. We repeated this process in kindergarten classrooms participating in the study in the fall of 2017 and enrolled 78% of students in participating classrooms who had not attended the public school BPS PreK program during the 2016–2017 academic year.

Direct Assessments

We trained research staff to reliability and then collected direct assessments of academic skills in the fall of 2017 (September 22nd through December 7th) and spring of 2018 (April 1st through June 15th), when study students were enrolled in kindergarten. We collected assessments for 75% of the student sample by October 22nd (within the first 5 weeks of school). We used the Prelanguage Assessment Scale (preLAS; Duncan & DeAvila, 1998) Simon Says and Art Show tests to determine the administration language for a subset of assessments (Barrueco et al., 2012). The preLAS assesses preliteracy skills and proficiency in English. Of the 462 children in the current study sample, 15 did not pass the preLAS and completed a subset of assessments in Spanish in the fall and four students did not pass the preLAS and completed assessments in Spanish in the spring.

Classroom Videotaped Observations

Prior to conducting observations, we reviewed all teachers' weekly schedules and identified a two- to three-hour block of instructional time which included substantial focus on language/literacy and another two- to three-hour block with some focus on math instruction. We worked with teachers to schedule observations during these two blocks of time when we would ideally observe a broad spectrum of curricular components. The length of videotaped observations did vary depending on the teacher's schedule and the activities happening at the school on the day of each recording. We observed kindergarten classrooms across two school days in late Winter/early Spring 2018 (98% of classrooms were observed in February and March; observation timing ranged from January 16th—April 5th; mean 6.47 days [$SD = 5.62$] between observations). We observed classrooms, on average, for a total time of 3.73 hr ($SD = .64$), ranging from 2.30 to 4.98 hr.

We used two video cameras during each observation session. One camera focused primarily on the teacher (and the teacher's microphone), and the other camera focused on the students in the classroom enrolled in the study to capture as much of the classroom activity as possible. A research assistant also took extensive notes when a child was off camera for any reason, including the reason why and what the child was engaged in. Before coding, we synchronized videos from the two observations to effectively track both the teacher and students as they moved between camera angles. We used Noldus Observer XT 13 software for coding videotapes using the Individualizing Student Instruction (ISI) observation measures. Coders followed and coded the experiences of each participating child. For any focal child who was off-camera, the coder relied on the research assistant's notes to capture the child's experiences.

Coders participated in multiple training sessions on ISI measures and were tested on the mastery of the codebook before coding. After training, coders had to show reliability on the ISI via coding four 20-minute video segments. Compared with a master-coded file, all coders scored $>.80$ Kappa on each of the four videos. Throughout the coding process, to prevent drift in reliability, we randomly selected and double-coded 20% of the video observations. After each round of double coding (five total rounds), coders discussed any coding disagreements. We calculated reliability in the Noldus Observer XT software which compares the duration of time of each code and the order/sequence of codes within a 15-s grace window. The average Kappa in kindergarten was .73, a similar level as past studies (e.g., average of .76 in Connor et al., 2009).

We also coded the teacher-focused videotapes from our observations using the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) which comprises Emotional Support, Classroom Organization, and Instructional Support domains. All coders participated in a 2-day training to learn the CLASS measure and then established reliability on a set of master codes created by the test developers. Coding of each videotape started once the instructional time began. As recommended by the measure's protocol (Pianta et al., 2008), coders used cycles of 20 minutes for observing and 10 minutes for scoring, which they repeated up to four times for each observation. Coding began when instruction commenced in the video and ceased after 80 minutes of observed time. We averaged scores across the four segments to calculate observation-specific scores and then we averaged scores across observations to generate one overall score for each classroom. We double-coded 20% of the observations to assess interrater reliability. The final ICCs representing interrater reliability (within 1 point) for the three domains were 96% for Emotional Support, 94% for Classroom Organization, and 88% for Instructional Support. We also conducted drift checks wherein coders had to code a master tape every three weeks to ensure they were still reliable.

Teacher Surveys

In the spring of the PreK and kindergarten years we asked teachers to complete a survey reporting on their demographic characteristics, teaching experience, and instruction. We used reports

on instruction to assess the degree to which children were exposed to advanced versus more basic content in kindergarten (described in further detail below). We used the data on demographics and teaching experiences to describe the sample.

Parent Surveys

In the fall of 2016 and 2017, using text messages and emails we contacted the consenting parents of all students who were selected for the study sample to complete a twenty-minute survey. 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, 94% of students in the current study had a parent who completed the survey in 2016 or 2017, allowing for fairly complete parent-reported covariate data.

Administrative Data From the School District

We accessed administrative records from the BPS district on students' demographic characteristics, history of enrollment in the BPS PreK 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.

Measures

Individualizing Student Instruction

To capture child-level learning experiences, we used the ISI Coding System (Connor et al., 2009). The ISI measures three sources of within-classroom variation in children's experiences: the amount of time a child is engaged in learning activities (vs. in management/routines or off task); the amount of time a child is exposed to different content areas (language/literacy, math, etc.); types of instruction (e.g., constrained vs. unconstrained activities; see Table S2 for full list); and the amount of time the child spends in different learning contexts (whole group, small group, individual, etc.). The current study uses the codes that capture the amount of time spent in different activity settings, specifically whole groups, small groups, and centers; the amount of overall time in language/literacy and math; and the amount of time spent in the language/literacy and math domains, further differentiated between constrained and unconstrained instruction.

The ISI measure coded in the current study relied heavily on a revised version that included math instructional codes and more nuanced noninstructional codes. We consulted with a math expert Dr. Deborah Ball to define these codes and checked reliability on them in the training and coding processes just as we did with all other ISI codes. For each target child, research assistants trained to reliability coded each second of observed time, switching codes as necessary to capture children's focal activities, content, and settings.

Table S2 in the online supplemental material provides breakouts of our overarching language/literacy and math constructs. Within literacy, for example, we coded the focal content of each activity each child engaged in using codes such as phonological awareness, fluency, and text reading. In math, we did the same, with codes such as number sense and concepts, operations, and data analysis. Following McCormick et al. (2021), Snow and Matthews (2016), and Paris (2005), we further grouped language and literacy and math codes into both total codes and into more constrained versus less constrained skills. To determine which ISI codes were unconstrained versus constrained, four of the current study's authors independently categorized each ISI code. Of 79 codes, all four coders agreed on the classification of 83% (65 codes) and at least one coder disagreed on 17% (14 codes). We asked one of the original authors of the theory to categorize these codes. Through this process, we reached agreement on the categorization of all 79 codes into unconstrained or constrained.

For the students with two observations (85% in kindergarten), we summed their data across both observation days to create aggregate child-level measures. We then conducted our analysis in terms of number of minutes in the domains of interest at the child-level (Connor et al., 2006), as well as percentage of time adjusting for total number of minutes observed (described in the robustness check section). To account for the students who only had one observation (15%), we included a dichotomous indicator in our key models (see analytic section).

Global Classroom Quality

We measured global classroom process quality using the Classroom Assessment Scoring System (CLASS) PreK (Pianta et al., 2008). CLASS measures three domains of teacher-child interactions: Emotional Support, Classroom Organization, and Instructional Support. All dimensions are directly scored on a 7-point scale, where a score of 7 represents high-quality except for negative climate which is reverse-coded. The CLASS and these three constructs show good psychometric reliability and validity in the literature, and prior studies examining associations between quality and children's outcomes have used this same three-factor structure (Burchinal, 2018). Prior work Guerrero-Rosada et al. (2021) assessed the psychometric properties of the CLASS three-factor model and found adequate fit to the data and good internal consistency.

Advanced Instructional Content

In the spring of each year, teachers reported on their math and literacy practices using items drawn from the Common Core Standards for English Language Arts and Math, ECLS-K Spring Classroom Teacher Questionnaire, and items originally created for this study. The team adapted 47% of items from the Common Core Standards (44% of language/literacy items, 50% of math items), drew 24% of items from the ECLS-K Spring Classroom Teacher Questionnaire (28% of language/literacy items, 20% of math items), and drafted the remaining 29% of items (28% Language/literacy, 30% math). Teachers reported on whether they taught each instructional practice, whether each practice was a major focus of instruction during the year, and whether it was taught to all or selected students.

We then coded how advanced these practices were, based on the expected grade level of the activities in the Common Core Standards for English Language Arts and Math and the Massachusetts State Standards, which includes PreK standards. Approximately 30% of practices were expected for the corresponding grade level (28% in the Pre-K survey and 32% in the kindergarten survey). To understand whether children were taught content below their grade level in kindergarten, we kept some of the expected PreK practices (17% across language/literacy and math)

in the kindergarten survey. All remaining practices (72% in PreK, 51% in kindergarten) were aligned with first or second grade. We assigned items at the corresponding grade level a score of 0, items one year ahead a score of 1, and items two years ahead a score of 2. We coded items expected at PreK as −1. We then calculated the average grade level taught as a major focus of instruction throughout the year, for each domain, in kindergarten. A score of 0 would indicate that practices were in line with grade level standards, while a positive score indicates that practices on average, were above the expected grade level.

Experience in Early Childhood Education

We used administrative data from the BPS district to identify students who were enrolled in the BPS PreK program implemented in public schools in 2016–2017. We also asked parents to report where their child spent the majority of his or her time during a regular week during the 2016–2017 year. Children who did not enroll in BPS PreK in 2016–2017 had a mix of experiences. They enrolled in Head Start, private child care center, a CBO PreK program implementing the BPS PreK curriculum, or did not attend any 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 284 children who attended BPS PreK during the 2016–2017 year and 178 who did not. All children, however, were enrolled in the same sets of kindergarten classrooms during the 2017–2018 year. Within the group of students who did not enroll in BPS PreK during the 2016–2017 year, 19% attended a CBO implementing the BPS PreK curriculum during their 4-year-old year, 20% attended a private child care center during their 4-year-old year, 23% attended a Head Start program during their 4-year-old year, and 38% did not enroll in formal care. We confirmed these categorizations through online searches of the reported locations coupled with follow-up phone calls to centers. We used district data to further clean this variable, describing a child as being a BPS PreK attender if he or she had been enrolled for 89 days or more (at least half the year). We then dummy coded these variables by assigning a 1 to a public BPS PreK attender and a 0 otherwise.

Language Skills

We used the Peabody Picture Vocabulary test IV (PPVT IV; Dunn & Dunn, 2007) to directly assess children's receptive language skills in the fall and spring of kindergarten. The PPVT IV is a nationally normed measure that has been used widely in diverse samples of 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. We assessed all children on the PPVT—regardless of whether they passed the PreLAS language screener—to describe an equivalent measure of language skills in English across the sample. As noted above, the proportion of students who did not pass the preLAS was small.

Literacy Skills

We used the Letter Naming Fluency (LNF) subtest from the teacher-reported Dynamic Indicators of Basic Literacy Skills–Next (DIBELS; Cummings et al., 2011) to measure children's literacy skills. This subtest has good reliability and good concurrent, predictive, and discriminant validity properties, is widely used, and is sensitive to intervention effects (e.g., Biancarosa et al., 2010). Although there were other DIBELS subtests available, we chose to use only the LNF subtest for the current study because it was the only one collected in both the fall and the spring of the kindergarten year. This decision also restricted the number of statistical tests and reduced concerns about multiple comparisons.

Math Skills

We used two different assessments to capture math skills. First, we used the Woodcock Johnson Applied Problems III (Woodcock et al., 2003) subtest to assess 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 et al., 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., 2003) and it has been used with diverse populations (Wong et al., 2008). We present results using the raw score, with English and Spanish assessments combined together.

The Woodcock-Johnson Applied Problems subtest has been criticized by some math experts because it is not particularly sensitive in the early childhood years, skips quickly to difficult items, and does not include geometry (Weiland et al., 2012). As such, we also used the Research-based Early Mathematics Assessment (REMA; Clements et al., 2008) to assess math skills in the fall and spring of kindergarten. 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 and colleagues (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, construction of shapes, spatial imagery, and measurement and patterning as being more unconstrained. 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. 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 had to achieve 90% or higher on a mock administration. The alpha reliabilities for the total scores (referred to as t scores) 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).

Child Characteristics From Administrative Data Used to Create Covariates

We used administrative data 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 free- or reduced-price lunch (FRPL; 1 if eligible; 0 if not) and gender (1 = female; 0 = not female). We set a dummy variable for Dual Language Learner (DLL) equal to 1 if the parent reported 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.

Family Characteristics

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 PreK. 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 GED or less, some college or two-year degree, and four-year degree (higher than a four-year degree as reference). All covariates used in the study have been shown to predict children's outcomes in other work (Choi et al., 2019; Powell et al., 2010; Reardon & Portilla, 2016), and are the same covariates used in our other published studies with these data (e.g., Guerrero-Rosada et al., 2021; McCormick et al., 2021).

Analytic Approach

Missing Data

Within the current study sample, there was a relatively low amount of missing data. All students had complete data on child-level information provided by the school district. Missingness on the main predictors of interest ranged from 7% to 16%. Up to 12% of students were missing parent survey covariate data, 2% were missing some outcome data on one of the assessments our research team collected, and up to 20% of students were missing information on the DIBELS data that was collected by the district. Because these data suggest that missingness was dependent on a particular set of demographic characteristics, we had some evidence to indicate that data were systematically missing. Little's MCAR test further suggested that the data were not Missing Completely at Random (χ^2 = 1193.649, $p < .01$). We made the assumption that the data were Missing at Random (MAR; Little & Rubin, 1989) and used multiple imputation to handle missing data of independent variables. First, we imputed 100 data sets using multivariate normal regression. We did not impute outcome variables. Results of these analyses are presented in the main text of the article.

Descriptive Analysis

To describe the sample before examining our research questions—and to consider potential selection into different kindergarten experiences—we first computed descriptive statistics on demographic variables and assessment scores in the fall and spring of kindergarten. We then examined individual-level instructional experiences, global classroom quality, and exposure to advanced instruction by enrollment in BPS PreK. We calculated zero-order correlations between the binary indicator for enrollment in BPS PreK, children's skills assessed in the fall and spring of the kindergarten year, and children's instructional experiences as measured on the ISI, the CLASS, and the teacher survey of advanced instructional content.

Multilevel Modeling to Examine Study Research Questions

We then used multilevel modeling to examine whether associations between enrollment in BPS PreK—relative to enrollment in any other type of center-based care and/or care at home—and gains in children's academic skills during the kindergarten year varied by individual instructional experiences (student-level), global classroom quality (classroom-level), and exposure to advanced instructional content (classroom-level). Examination of intraclass correlations (ICCs; conducted before running predictive models) at the classroom- and school-levels revealed that between-group variation in outcomes was non-negligible at both levels (range from .03–.31 at the classroom-level; range from .03–.15 at the school-level), indicating that a three-level model (Level 1: students; Level 2: classrooms; Level 3: schools) was the most appropriate fit to the data across outcomes (Raudenbush & Bryk, 2002). We then used slightly different approaches (explained below) to model the theorized sustaining environments depending on whether they were measured at Level 1 (individual instruction experiences) or Level 2 (global classroom quality and advanced instructional content).

Individual Instructional Experiences

We grouped individual instructional experiences into four different conceptual blocks—time in instruction versus noninstruction (i.e., off-task or in management/routines), time spent in different activity settings (whole groups, small groups, centers, individual time), time spent in constrained and unconstrained language/literacy and math instruction, and time spent in any type of language/literacy and math instruction. We fit separate models for each conceptual block, including the individual variables within each conceptual block as predictors in the same model (see Equation 1 below). This approach allowed us to balance the need to be parsimonious and limit the number of models, while also wanting to align the models with the specific sets of factors theorized to be important for sustaining the PreK boost. Examination of bivariate correlations (Table S3 in the online supplemental material) showed that the factors within blocks had low to moderate correlations (r = .02 to r = .61), allowing us to adjust for conceptually similar factors in the same model that could bias results.

Prior to fitting the models needed to answer our research questions, we first we regressed each outcome in the spring of kindergarten on the indicator for assignment to BPS PreK and a conceptual block of variables representing different instructional

experiences (with separate models for the conceptual blocks of instruction vs. noninstruction, constrained and unconstrained instruction, overall math and language/literacy, and activity settings), adjusting for a host of controls at the student-level, including the level of the outcome measured in the fall of kindergarten. This model allowed us to understand associations between enrollment in BPS PreK and the instructional experiences of interest on gains in students' skills, without taking into account how those associations might vary by enrollment in BPS PreK. We then added *interactions* between the BPS PreK indicator and each of the individual instructional experiences for each conceptual block of interest to this baseline model. We grand mean centered all continuous variables (Cohen et al., 2013). The model for instruction versus noninstruction as a conceptual block with interactions is displayed here as an example:

$$Y_{ijk} = \beta_0 + \beta_1 BPSPreK_{ijk} + \beta_2 InstructionalTime_{ijk} + \beta_3 Offtask_{ijk} + \beta_4 ManageRoutine_{ijk} + \beta_5 BPSPreKxInstructionalTime_{ijk} + \beta_6 BPSPreKxOfftask_{ijk} + \beta_7 BPSPreKxManageRoutine_{ijk} + \gamma_{ijk} + \mu_{jk} + \zeta_k + \varepsilon_{ijk}, \quad (1)$$

where i denotes students and j and k represent kindergarten classrooms and schools, respectively, Y is the child-level outcome measure, BPSPrek is a dichotomous indicator set to one if student i attended the BPS PreK program and 0 otherwise. InstructionalTime, Offtask, and ManageRoutine represent the time spent in instructional time, off-task, and in management/routines for student i in classroom j and school k, grand mean centered. There are three interaction terms between the indicator for BPS PreK attendance and each instructional experience within the conceptual block. γ is a vector of student-level covariates including the level of the outcome in the fall of PreK,[2] μ is the classroom-level random intercept, ζ is a school-level random intercept, and ε is the child-level error term. Statistically significant interactions in this model—the coefficients for $\beta 5$, $\beta 6$, and $\beta 7$—would demonstrate evidence that the association between enrollment in public BPS PreK and gains in children's skills in kindergarten vary by that particular instructional experience, adjusting for the other experiences. We then fit additional models for the other three conceptual blocks of instructional experiences (activity settings, time in constrained/unconstrained instruction, time in instructional domains) and repeated the process for each outcome. We fit 16 total models to answer the first research question (four outcomes × four conceptual blocks). Importantly, all of our multilevel models made the explicit assumption that the independent variables were not correlated with unobserved variables. However, given descriptive findings—presented below—that students' demographic characteristics included as covariates were associated with both BPS PreK enrollment and some of the instructional experiences of interest, this assumption is very unlikely to hold. As such, it is important to interpret all coefficients as associations, adjusting for covariates, rather than interpreting any of the results causally.

Global classroom quality and exposure to advanced instructional content. We adjusted our modeling approach in Equation 1 to test the remaining theorized sustaining environments, both measured at the classroom level. We first fit models without interactions as we did above and then included interactions between the classroom-level variables and the indicator for BPS PreK enrollment to answer our main research questions. The equation for global classroom quality—fit using the CLASS emotional support domain in kindergarten—illustrates this approach:

$$Y_{ijk} = \beta_0 + \beta_1 PublicPreK_{ijk} + \beta_2 CLASS\_ES_{jk} + \beta_3 PublicPreKxCLASS\_ES_{ijk} + \gamma_{ijk} + \mu_{jk} + \zeta_k + \varepsilon_{ijk}. \quad (2)$$

The sustaining environment here is measured at the classroom level and the coefficient of interest ($\beta 3$) is on a cross-level interaction. The vector of covariates does not include variables related to the ISI measure. In addition, to align with prior work (e.g., Carr et al., 2019), we decided to use three separate models to test each of the CLASS domains separately and then we fit a fourth model with the CLASS total score as an additional check. We used the same approach to examine advanced instructional content as a sustaining environment by fitting separate models for grade level of math instruction and grade level of language/literacy instruction and interactions with the indicator for enrollment in BPS PreK. Each of these final models included main effects for grade level of math and language/literacy instruction, as they were moderately correlated ($r = .39$).

This study was not preregistered. Data are not yet publicly available but will be shared as a restricted access file—available upon application—after the conclusion of the grant period in 2023.

Results

Descriptive and Correlational Analysis

We found a number of differences in the demographic characteristics of children when comparing children in the BPS PreK and comparison groups. As illustrated in Table 1, children in the BPS PreK sample were less likely than children in the comparison group to be Black or eligible for free- or reduced-price lunch. They were more likely to be White, have parents who were married, and have a parent with more education than a four-year college degree. These racial/ethnic and socioeconomic disparities in enrollment in BPS public PreK mirror prior work done with district-wide data collected in 2009 and 2010 (Shapiro et al., 2019). Despite the majority of students in both groups being eligible for free- or reduced-price lunch, the magnitude of these differences indicates that, taken together, the BPS PreK attenders were more advantaged than students in the comparison group. Importantly as well, Black students are underrepresented in BPS PreK when compared with the population of students who eventually enroll in BPS public elementary schools (and kindergarten in particular). A complementary study done by our team has found that Black

[2] Covariates for this model are the level of the outcome in the fall of kindergarten, eligibility for free/reduced price lunch, female, DLL, race/ethnicity, age, mother's age at first birth, household size, parental employment, parental marital status, parent age, parent education, total number of ISI observations (1 or 2), the total amount of time that the child was observed on the ISI, and the amount of time between the fall and spring kindergarten assessments.

families live further away from BPS PreK public school programs, a potential mechanism explaining this disparity (Weiland et al., 2021). There may be other factors affecting Black families' access to the public PreK program as well, but those have been less studied to date. These selection issues are important to remember when interpreting the results of our future predictive models. Although models control for a robust set of covariates to address this selection, it is impossible to account for all unobservable characteristics likely associated with these observable differences.

Table 2 summarizes students' scores on assessments of literacy, language, and math, the amount of time they spent in various types of instruction, and their global classroom quality and exposure to advanced instruction content during kindergarten. As expected, children who attended BPS PreK had higher scores, on average, than nonattenders on language (PPVT), literacy (DIBELS), and math (REMA & WJAP) skills in the fall and spring of kindergarten, but the group-based difference on the DIBELS was no longer statistically significant by the spring of kindergarten. Examination of standardized scores on the PPVT and the WJAP suggests that students in the current sample, on average, scored higher than the national average for five-year-old children.

When we examined individual learning experiences in kindergarten (see Table 2), we found that students spent a substantial amount of time in instruction (compared with off-task or in management/routines), spent much more time in language/literacy instruction and activities than math, and spent more time in whole groups compared with small groups, centers, and individual time. Most of children's time in math instruction was constrained but language/literacy instruction was more balanced between constrained and unconstrained activities. We then considered how these instructional experiences varied for children in the group of BPS PreK attenders relative to the comparison group. Group-based differences emerged with respect to activity setting. Specifically, children who attended BPS PreK spent about 7 more minutes in centers in kindergarten, $t(383) = -1.97$, $p < .05$, and 8 more minutes in individual work, $t(383) = -2.32$, $p < .05$, compared with nonattenders. There was a trend-level difference ($p < .10$) for time spent in whole group; BPS PreK attenders spent about 8 fewer minutes in whole group than nonattenders, $t(383) = -1.92$, $p = .06$. None of the remaining comparisons was statistically significant.

Compared with children who did not attend the BPS PreK program, children who did experienced higher levels of classroom process quality on all three domains of the CLASS (instructional support: mean difference = .17; $t[400] = -3.05$, $p < .01$; emotional support: mean difference = .14; $t[400] = -2.82$, $p < .01$; classroom organization: mean difference = .16, $t[400] = -2.48$, $p < .05$)), as well as the total score (mean difference = .16; $t[400] = -3.20$, $p < .01$). Average grade level of language/literacy practices was slightly higher for students who attended BPS PreK (mean difference = .02; $t[429] = -2.0$, $p < .05$) but there was no difference between groups on the average level of math practices. Taken together, these findings revealed that enrollment in the BPS PreK program in our sample was associated with access to higher-quality learning and instructional experiences in about half of the explored indicators. It is critical to interpret the results of predictive models taking these group-based differences into account.

Work from the correlational analysis (Table S3) showed that enrollment in BPS PreK was moderately associated with all the outcomes in the spring of kindergarten, except the DIBELS. Examination of the instructional indicators revealed moderate correlations between variables within conceptual blocks, suggesting that our decision to fit models that controlled for instructional experiences within similar domains was appropriate.

Individual Instructional Experiences as Sustaining Environments

In Table 3 we summarize associations between enrollment in BPS PreK and students' gains in language, literacy, and math skills during kindergarten as well as associations between students' instructional experiences in kindergarten and gains in skills, not taking into account the interaction of those factors. In line with our team's prior work, across all models, we found consistent evidence of a statistically significant, negative association between BPS PreK and gains in literacy skills suggesting full fadeout/convergence for literacy skills, on average. However, as reported in prior work (McCormick et al., 2021), the coefficients for BPS PreK on the other three outcomes were negative but not statistically significant, suggesting the maintenance of the PreK boost through the spring of kindergarten.

With respect to main effects of the theorized sustaining environments, there was evidence that more time spent in constrained literacy instruction ($\gamma = -.09$, $SE = .04$, $p < .05$), constrained math instruction ($\gamma = -.14$, $SE = .07$, $p < .10$), and direct language/literacy instruction ($\gamma = -.07$, $SE = .03$, $p < .05$) were associated with fewer gains in language skills. In other models there was evidence that more time spent in centers ($\gamma = -.16$, $SE = .08$, $p < .05$) and small groups ($\gamma = -.12$, $SE = .07$, $p < .10$) was associated with fewer gains in literacy skills (but only literacy skills) on average, adjusting for time in other activity settings. With respect to math, we only found that time spent off task was associated with fewer gains on the REMA ($\gamma = -.04$, $SE = .02$, $p < .05$), adjusting for time spent in instruction and in management/routines. More in-depth discussion of these main effects can be found in complementary work done by our team examining these ISI data in detail (Weiland et al., 2021).

The next set of results, summarized in Table 4, reports on how enrollment in BPS PreK and gains in skills during kindergarten varied by these learning experiences in kindergarten. We found mostly null results when interacting features of children's individual instructional experiences and their enrollment in BPS PreK to predict gains in their language, literacy, and math scores in kindergarten.[3] There were, however, eight statistically significant interactions at $p < .05$ and four interactions at $p < .10$. We explain those findings below. We plotted statistically significant interactions to aid in interpretation and show only two of these in the main text. See Figures S1–S5 for additional plots for interactions that were generally robust across sensitivity checks in terms of magnitude and statistical significance.

Language and Literacy Outcomes

As shown in Table 4, in our models predicting gains in language skills, there was a positive interaction between enrollment in BPS

[3] Table 3 summarizes the interaction coefficients of interest across all models.

Table 2
Descriptive Statistics for Study Variables, by Enrollment in BPS Prekindergarten Program

	BPS prekindergarten				Other care							
	Fall of K		Winter/Spring of K[a]		Fall of K		Winter/Spring of K[a]		Comparison of means			
Variable of interest	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Fall of K	Stat. Sig.	Winter/Spring of K	Stat. Sig.
PPVT raw score	94.1	26.1	107.24	25.38	78	27.48	91.83	26.41	16.10	***	15.41	***
Woodcock Johnson Applied Problems	17.11	5.12	20.15	4.68	14.11	4.95	17.9	4.28	3.00	***	2.25	***
REMA t-score	41.32	5.5	44.46	5.18	38.09	5.52	42.13	4.56	3.23	***	2.33	***
DIBELS letter naming fluency score	27.06	16.81	51.06	18.63	17.65	15.18	49.13	16.88	9.41	***	1.93	
Individualizing Student Instruction (ISI) measures												
Time on instruction vs. other												
Total instruction	—	—	164.84	48.84	—	—	162.05	54.13			2.79	
Off-task	—	—	14.26	12.33	—	—	16.53	14.6			−2.27	
Management/routines	—	—	49.19	22.18	—	—	50.18	23.36			−0.99	
Primary content of instruction												
Language/literacy												
Total	—	—	93.17	35.77	—	—	97.69	43.57			−4.52	
Constrained	—	—	36.88	21.85	—	—	39.04	22.47			−2.16	
Unconstrained	—	—	42.29	25.71	—	—	39.79	24.44			2.50	
Math												
Total	—	—	32.21	14.75	—	—	29.41	16.30			2.80	†
Constrained	—	—	20.81	13.14	—	—	18.99	12.71			1.82	
Unconstrained	—	—	3.82	8.27	—	—	3.32	9.22			0.50	
Close-ended	—	—	8.84	9.10	—	—	7.59	7.47			1.25	
Open-ended	—	—	15.84	11.02	—	—	14.63	11.63			1.21	
Activity setting												
Centers	—	—	45.94	35.70	—	—	38.75	33.11			7.19	*
Small groups	—	—	21.00	20.70	—	—	24.10	23.77			−3.10	
Whole groups	—	—	100.71	35.16	—	—	108.44	43.20			−7.73	†
Individual time	—	—	55.6	36.04	—	—	47.74	25.14			7.86	*
Classroom process quality												
CLASS total score	—	—	4.86	0.51	—	—	4.70	0.42	—	—	0.16	**
Classroom emotional support	—	—	5.8	0.5	—	—	5.66	0.47	—	—	0.14	**
Classroom organization	—	—	5.89	0.63	—	—	5.73	0.59	—	—	0.16	*
Classroom instructional support	—	—	2.56	0.64	—	—	2.39	0.49	—	—	0.17	**
Exposure to advanced instructional content												
Grade level of language/literacy practices	—	—	0.46	0.14	—	—	0.44	0.12	—	—	0.02	*
Grade level of math practices	—	—	0.03	0.26	—	—	0.05	0.3	—	—	−0.02	

Note. BPS = Boston Public Schools; PPVT = Peabody Picture Vocabulary test; REMA = Research-based Early Mathematics Assessment; DIBELS = Dynamic Indicators of Basic Literacy Skills; Stat. sig. = Statistical significance. BPS PreK, $N = 284$, Other care, $N = 178$.
[a] The ISI measure was collected in the late winter/early spring of the kindergarten year while spring assessments were collected in April-June.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$,

Table 3
Results of Multilevel Models Examining Associations Between BPS PreK, Individual Instructional Experiences and Gains in Language/Literacy and Math Skills

Variable of interest	Language (PPVT)		Literacy (DIBELS)		Math (WJAP)		Math (REMA)	
	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*
Time on instruction vs. other								
BPS PreK	−1.79	1.63	−5.31***	1.79	−0.35	0.35	−0.49	0.35
Total instruction	−0.05	0.05	−0.09†	0.05	0.00	0.01	0.00	0.01
Off-task	−0.10	0.07	−0.04	0.08	−0.02	0.01	−0.04*	0.02
Management/routines	−0.06	0.05	−0.08	0.06	−0.01	0.01	0.00	0.01
Time in constrained and unconstrained instruction								
BPS PreK	−1.40	1.61	−4.82**	1.79	−0.33	0.35	−0.48	0.35
Constrained literacy	−0.09*	0.04	0.03	0.05	0.00	0.01	0.00	0.01
Unconstrained language	−0.05	0.03	0.00	0.04	0.00	0.01	0.00	0.01
Constrained math	−0.14†	0.07	−0.06	0.09	0.01	0.02	0.00	0.02
Unconstrained math	−0.07	0.09	−0.05	0.12	0.00	0.02	0.01	0.02
Total time in language/literacy and math instruction								
BPS PreK	−1.66	1.61	−4.92**	1.80	−0.34	0.35	−0.46	0.35
Total language/literacy	−0.07*	0.03	0.00	0.03	0.00	0.01	0.00	0.01
Total math instruction	−0.08	0.06	0.00	0.07	0.01	0.01	0.00	0.01
Activity setting								
BPS PreK	−1.67	1.62	−5.23**	1.79	−0.38	0.35	−0.50	0.35
Centers	0.06	0.06	−0.16*	0.08	0.01	0.01	0.01	0.02
Small groups	−0.02	0.06	−0.12†	0.07	−0.01	0.01	0.02	0.01
Whole groups	−0.01	0.06	−0.12	0.08	0.00	0.01	0.00	0.02
Individual time	0.02	0.06	−0.12	0.07	−0.01	0.01	0.01	0.01
Classroom process quality								
BPS PreK	−1.36	1.02	−4.98**	1.76	−0.24	0.27	−0.45	0.32
Instructional support	−3.25	2.43	−2.92	3.25	−0.26	0.52	0.02	0.57
Emotional support	3.22	2.55	−8.24**	3.38	1.12*	0.53	0.73	0.58
Classroom organization	1.87	2.03	−5.40*	2.76	0.99*	0.42	0.97*	0.47
CLASS total score	1.30	2.86	−8.33**	3.73	1.07†	0.61	0.92	0.67
Exposure to advanced instructional content								
BPS PreK	−1.53	1.13	−5.02**	1.54	−0.31	0.25	−0.42	0.37
Grade level of language/literacy	−1.52	10.27	−1.70	13.22	−3.95†	2.28	2.19	2.67
Grade level of math	−4.74	4.71	−15.15*	6.28	0.16	1.06	−1.66	1.28

Note. PreK = prekindergarten; BPS = Boston Public Schools; PPVT = Peabody Picture Vocabulary test; REMA = Research-based Early Mathematics Assessment; DIBELS = Dynamic Indicators of Basic Literacy Skills; WJAP = Woodcock Johnson Applied Problems. Models testing the ISI variables adjust for the effects of the other ISI variables—and the other interactions —in each conceptual block. Models testing the CLASS and grade level of instruction variables do not adjust for the other predictor(s) in that block in the main analysis (but do so in the robustness checks; see Table S1).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

PreK and time spent in unconstrained language interaction that was statistically significant at $p < .05$ (γ = .11, *SE* = .05, $p < .05$). For students in the comparison group (see full results including main effects in the presence of the interactions in Tables S4 and S5), there was a negative and statistically significant association between unconstrained language instruction and gains in children's language skills (γ = −.14, *SE* = .04, $p < .01$). In contrast, this association appeared to be null for BPS PreK attenders. Thus, as illustrated in Figure S1, unconstrained language instruction was associated with PreK persistence in language skills. We also found a negative and statistically significant interaction between enrollment in BPS PreK and time spent in constrained math instruction (γ = −.28, *SE* = .12, $p < .05$) predicting gains in language skills. For non-BPS PreK attenders there was a null association between time spent in constrained math instruction and gains in language skills (γ = .02, *SE* = .10, $p < .84$; see Table S4). Yet, as illustrated in Figure S2, the initial benefit of BPS PreK on language skills was maintained through the spring for children who spent less time in constrained math instruction. Relatedly, there was also a negative interaction between time spent in math and enrollment in BPS PreK (γ = −.18, *SE* = .09, $p < .05$). However, given that the large majority of time spent in math was constrained (∼87%) these results may be somewhat redundant with one another. Taken together, they demonstrate that constrained math instruction in kindergarten was associated with PreK fadeout/convergence whereas less time in constrained math instruction was associated with persistence in language skills.

When examining gains in literacy skills, there was a positive and statistically significant interaction between enrollment in BPS PreK and time spent in unconstrained language instruction (γ = .17, *SE* = .07, $p < .05$). For non-BPS PreK attenders, the association between unconstrained language instruction and gains in literacy skills was negative and statistically significant (γ = .13, *SE* = .06, $p < .05$; see Table S5). In contrast, this association was positive for BPS PreK attenders. Thus, as illustrated in Figure 1, unconstrained language instruction in kindergarten was associated with PreK persistence in literacy skills. There was also a statistically significant and negative interaction between BPS PreK enrollment and time spent in constrained math (γ = −.32, *SE* = .16, $p < .05$) when predicting literacy skills. The association between constrained math instruction and gains in literacy skills for the non-BPS PreK attenders was positive

Table 4

Results of Multilevel Models Examining Variation in Associations Between BPS Prekindergarten and Gains in Language/Literacy and Math Skills by Individual Instructional Experiences in Kindergarten, Process and Instructional Quality, and Grade Level of Instruction

Interaction Effects	Language (PPVT)		Literacy (DIBELS)		Math (WJAP)		Math (REMA)	
	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*
Time on Instruction vs. Other × BPS PreK								
Total instruction	0.00	0.03	0.04	0.03	0.00	0.01	0.00	0.01
Off-task	0.03	0.12	0.07	0.13	−0.03	0.03	0.02	0.02
Management/routines	−0.02	0.07	−0.14	0.09	0.02	0.02	0.00	0.02
Time in Constrained and Unconstrained Instruction × BPS PreK								
Constrained literacy	0.00	0.07	0.05	0.08	0.00	0.01	−0.04**	0.02
Unconstrained language	0.11*	0.06	0.17*	0.07	−0.02	0.01	0.02	0.01
Constrained math	−0.28*	0.12	−0.32*	0.16	0.01	0.03	0.01	0.03
Unconstrained math	−0.18	0.17	0.13	0.19	−0.01	0.04	0.00	0.04
Total Time in Language/Literacy and Math Instruction × BPS PreK								
Total language/literacy	0.04	0.04	0.07†	0.04	−0.01	0.01	−0.01	0.01
Total math instruction	−0.18*	0.09	−0.14	0.11	0.00	0.02	0.01	0.02
Activity Setting × BPS PreK								
Centers	−0.04	0.05	−0.01	0.05	0.01	0.01	0.01	0.01
Small groups	0.08	0.07	−0.13†	0.08	−0.01	0.01	−0.02	0.02
Whole groups	−0.01	0.04	0.11*	0.05	0.00	0.01	0.01	0.01
Individual time	−0.07	0.06	−0.06	0.06	0.00	0.01	0.00	0.01
Classroom Process Quality × BPS PreK								
Instructional support	4.27	2.69	−1.74	3.45	−0.29	0.58	−0.15	0.60
Emotional support	−0.51	2.91	6.00†	3.54	−0.89	0.63	−0.04	0.64
Classroom organization	−1.01	2.38	4.54	2.91	−0.82	0.51	−0.66	0.52
CLASS total score	0.73	3.22	4.85	3.90	−1.14	0.70	−0.53	0.71
Exposure to Advanced Instructional Content × BPS PreK								
Grade level of language/literacy	0.72	10.31	14.25	12.99	4.21†	2.31	−0.46	2.55
Grade level of math	6.09	4.40	13.23**	5.10	0.92	1.00	1.51	1.07

Note. PreK = prekindergarten; BPS = Boston Public Schools; PPVT = Peabody Picture Vocabulary test; REMA = Research-based Early Mathematics Assessment; DIBELS = Dynamic Indicators of Basic Literacy Skills; WJAP = Woodcock Johnson Applied Problems. Models testing the ISI variables adjust for the effects of the other ISI variables - and the other interactions—in each conceptual block. Models testing the CLASS and grade level of instruction variables do not adjust for the other predictor(s) in that block in the main analysis (but do so in the robustness checks; see online supplemental materials).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

but null (γ = .12, *SE* = .13, *p* = .36; Table S5). The PreK boost was more likely to persist when children spent less time in constrained math instruction (Figure S3).

With respect to format of instruction, we also found a statistically significant and positive interaction between enrollment in BPS PreK and time spent in whole group (γ = .11, *SE* = .05, $p < .05$). Examination of the nature of this interaction (see Figure S4) revealed that fadeout/convergence was more likely when children spent less time in whole group instruction in kindergarten. In contrast, more time in whole group in kindergarten was associated with PreK persistence in literacy skills, adjusting for the mean levels of time spent in small group, centers, and individual time. Importantly, however, both groups of children were predicted to do better by the spring when they spent less time in whole group. Interestingly, we also found a negative interaction between small group and BPS PreK enrollment at $p < .10$ (γ = −.13, *SE* = .08, $p < .10$) with plotted results suggesting that non-BPS PreK attenders had slower gains in literacy skills when they spent more time in small groups whereas gains in literacy skills for BPS PreK attenders did not vary by small group instruction. As such, small group instruction supported PreK persistence in literacy skills but this was driven by slower gains for non-BPS PreK attenders (rather than faster gains for attenders). There was another interaction at $p < .10$ between PreK enrollment and time spent in language/literacy (γ = .07, *SE* = .04, $p < .10$), suggesting that persistence in literacy skills was also more likely when children spent more time in language/literacy during kindergarten.

Math Outcomes

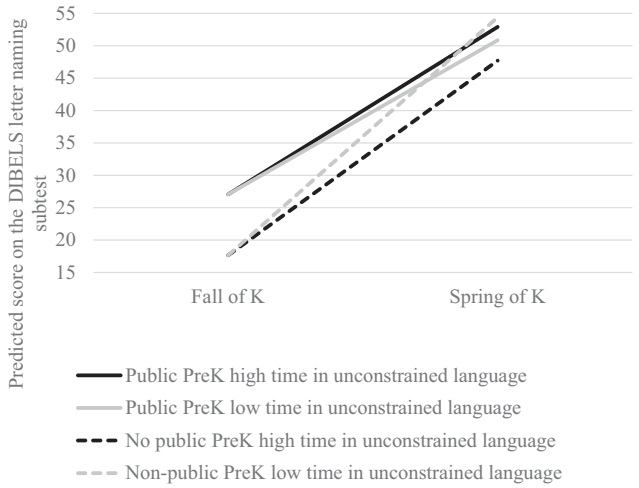
The bulk of our findings in the models examining math outcomes were null. However, in our models examining gains in math skills during kindergarten on the REMA measure, we found a statistically significant and negative interaction between BPS PreK attendance and time spent in constrained literacy instruction (γ = −.04, *SE* = .02, $p < .05$). Constrained literacy instruction was positively associated with non-BPS PreK attenders' gains in math skills (γ = .03, *SE* = .02, $p < .10$). Taken together, and illustrated in Figure 2, results demonstrated that constrained literacy instruction was associated PreK fadeout/convergence in math skills. There was no evidence that the other indicators of instruction affected associations between enrollment in BPS PreK and gains in children's math skills and there were no statistically significant interactions in ISI models predicting the WJAP.

Global Classroom Quality as a Sustaining Environment

As summarized in Table 3, in our initial models that did not include interactions there was evidence that higher classroom emotional support (γ = −8.24, *SE* = 3.38, $p < .01$) and classroom organization (γ = −5.40, *SE* = 2.76, $p < .05$) were associated with

Figure 1

Variation in Associations Between Enrollment in BPS PreK and Gains in Literacy Skills, for Students With High and Low Time Spent in Unconstrained Language Instruction in Kindergarten



Note. High time is plotted for one standard deviation above the mean and low time is plotted as one standard deviation below the mean. Models adjust for the level of the outcome in the fall of kindergarten, eligibility for free/reduced price lunch, female, DLL, race/ethnicity, age, mother's age at first birth, household size, parental employment, parental marital status, parent age, parent education, total number of Individualizing Student Instruction (ISI) observations (1 or 2), the total amount of time that the child was observed on the ISI, and the amount of time between the fall and spring kindergarten assessments. Predicted values are based on reference groups being equal to 0 and continuous variables at the mean level.PreK = prekindergarten; BPS = Boston Public Schools; DIBELS = Dynamic Indicators of Basic Literacy Skills.

fewer gains in literacy skills during kindergarten. In contrast, emotional support was associated with bigger gains in math on the WJAP ($\gamma = 1.12$, $SE = .53$, $p < .05$) and classroom organization was associated with larger gains in math on the WJAP ($\gamma = .99$, $SE = .42$, $p < .05$) and the REMA ($\gamma = .97$, $SE = .47$, $p < .05$) in kindergarten.

In the next series of models, there was a positive interaction at $p < .10$ between emotional support and the indicator for BPS PreK ($\gamma = 6.00$, $SE = 3.54$, $p = .09$) in the model predicting gains in literacy skills in kindergarten. PreK persistence in literacy skills was more likely when children were exposed to lower levels of emotional support in kindergarten. Emotional support on average, however, was quite high so classrooms 1 *SD* below the mean still had a score of 5.24 on a 1–7 scale. Other interactions between the CLASS domains and enrollment in BPS PreK were not statistically significant.

Exposure to Advanced Instructional Content as a Sustaining Environment

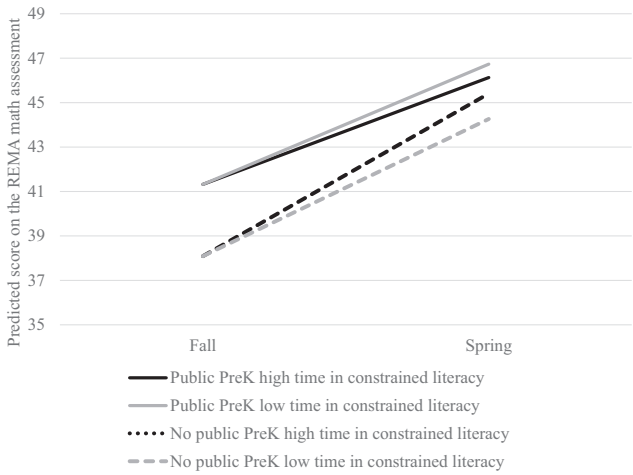
In models without interactions, we found that more advanced math content was actually associated with fewer gains in children's literacy skills ($\gamma = -15.15$, $SE = 6.28$, $p < .05$) and more advanced language/literacy instruction was associated with fewer gains in math skills as measured on the WJAP ($\gamma = -3.95$, $SE = 2.28$, $p < .10$). There was no other evidence that level of content predicted gains for other learning domains and outcomes. There was one positive and statistically significant interaction that emerged when we examined grade level of math instruction as a sustaining environment in our model predicting gains in literacy skills ($\gamma = 13.23$, $SE = 5.10$, $p < .01$). PreK persistence for literacy skills was more likely when kindergarten teachers reported more advanced math instruction. In contrast, convergence was more likely (as illustrated in Figure S5) when kindergarten teachers reported less advanced math instruction. There was also a trend-level positive interaction between BPS PreK and advanced language/literacy instruction in our model predicting gains in math skills, as measured using the WJAP ($\gamma = 4.21$, $SE = 2.31$, $p = .08$). The pattern was similar to our findings for literacy skills, such that persistence on math skills was more likely when kindergarten teachers reported more advanced language/literacy instruction. All other interactions between grade level of instruction and enrollment in BPS PreK were not statistically significant at the $p < .05$ or $p < .10$ levels.

Sensitivity Analyses

We conducted a number of robustness checks to examine the sensitivity of results to potential omitted variables, concerns about endogeneity, model specification (including considering models with school and classroom fixed effects and regressions with robust standard errors), and operationalization of the ISI variables.

Figure 2

Variation in Associations Between Enrollment in Public BPS PreK and Gains in Math Skills, for Students With High and Low Time Spent in Constrained Literacy Instruction in Kindergarten



Note. High time is plotted for one standard deviation above the mean and low time is plotted as one standard deviation below the mean. Models adjust for the level of the outcome in the fall of kindergarten, eligibility for free/reduced price lunch, female, DLL, race/ethnicity, age, mother's age at first birth, household size, parental employment, parental marital status, parent age, parent education, total number of Individualizing Student Instruction (ISI) observations (1 or 2), the total amount of time that the child was observed on the ISI, and the amount of time between the fall and spring kindergarten assessments. Predicted values are based on reference groups being equal to 0 and continuous variables at the mean level. PreK = prekindergarten; BPS = Boston Public Schools; REMA = Research-based Early Mathematics Assessment.

There were 12 interactions that were statistically significant at $p < .10$ in our main analyses (8 at $p < .05$ and 4 at $p < .10$). Six of the 12 interactions (highlighted in gray in Table S1) were robust in terms of magnitude and statistical significance across all checks. The interaction between BPS PreK enrollment and unconstrained language in the model predicting language skills remained similar in terms of magnitude but was no longer statistically significant in two of four ISI checks. The interaction between total time in language/literacy instruction and PreK—which was originally only significant at the $p < .10$ level—was consistent in magnitude but not robust in terms of statistical significance in three of four ISI checks. The interaction between PreK and small group changed in magnitude and statistical significance across checks, suggesting that it might have been spurious. The interaction between PreK and emotional support in the model predicting literacy skills and between PreK and grade level of language/literacy instruction in the model predicting math skills were not robust across checks in terms of magnitude or significance.

Discussion

This study aimed to use a range of nuanced, individual- and classroom-level measures of children's instructional experiences in kindergarten settings to build on existing work and further explore the sustaining environments hypothesis as an explanation for PreK persistence. Although the research on this topic has been quite mixed to date, our goal was to use a comprehensive measurement strategy operationalizing a broader range of sustaining environments than have been typically examined to address existing measurement limitations in this literature. As such, we examined 13 possible student-level instructional factors to explain PreK fadeout/convergence in children's language, literacy, and math skills, fit in models by four conceptual blocks. This approach is unique in that we were able to observe and operationalize these factors at the individual level. Accordingly, we were able to address concerns about within-classroom variation in children's experiences—potentially related to PreK attendance—as well as concerns about limited statistical power when using classroom-level measures as sustaining environments. Subsequent classroom-level work allowed us to also examine sustaining environments that had been tested in prior studies—global classroom process and instructional quality and advanced instructional content—facilitating comparison with extant work.

Yet, even with this robust, multilevel, and multimodal measurement approach we did not find clear results across outcomes on the key instructional experiences that consistently predicted PreK persistence. Taken as a whole, our study aligns with extant meta-analytic work finding limited consistent evidence that the quality of learning environments subsequent to PreK explains the fadeout/convergence phenomenon (Bailey, Duncan, et al., 2020). Indeed, the large majority of our findings were null with point estimates on the interactions close to 0, which maps onto meta-analytic research (Bailey, Jenkins, et al., 2020) finding a similar trend. We observed this pattern across models despite the relative socioeconomic advantage that the BPS PreK attenders had over the students in the comparison group. For example, although we might have expected to see stronger gains in kindergarten for BPS PreK attenders owing to this upward bias potentially affecting future skills and access to higher-quality sustaining environments, we did not find evidence to that effect.

In line with earlier work (Abenavoli, 2019; Pearman et al., 2020), there may be a complex combination of different factors—for example, the quality of PreK program, the counterfactual experiences of children who do not attend PreK, the peer learning environment, the global quality of the school environment—that cumulatively or interactively contribute to convergence. We do not yet know the specific combination of elements that explain PreK persistence, nor do we know if consistency across settings is a reasonable expectation given wide variation in instructional contexts, counterfactuals, and student demographics. Despite strong theoretical underpinnings, there continues to be limited evidence that the quality of children's learning environments subsequent to PreK—measured in a range of different ways—consistently explain PreK persistence across different types of outcomes.

We further concede that the current study may have faced some key methodological challenges that hold lessons for future work in this area. For example, even when using individual-level data on sustaining environments, the predictive analyses may have been underpowered to detect statistically significant interactions. In designing this study, we sought to collect a diverse set of reliable, valid, and nuanced measures that were very resource intensive to collect, particularly the individual child-level observations of student learning. Although we were able to collect these data for a sample of several hundred students in kindergarten, that sample size may still not have been sufficient to detect statistically significant interactions (Porter, 2018). Future research interested in building on this paper should prioritize addressing this power constraint and use our results and estimates to guide study design and goals related to student enrollment. For example, researchers could use the coefficients from our main effects and interaction terms to identify the sample size needed to detect statistically significant interactions for different percentiles of the coefficient of interest (e.g., half to three quarters of the size of the coefficient we detected; Kraft, 2020). By engaging in that exercise, future research can then determine whether such a sample size is really feasible for a prospective study given the high cost of collecting and coding individual-level observational data. Leveraging data across different research teams and data sets with similar sets of measures may also hold promise for addressing statistical power constraints in light of limited resources within individual research projects.

Constrained and Unconstrained Instruction in Kindergarten as Sustaining Environments

If one were to only focus on the statistically significant results from this study—and ignore the largely null pattern of findings—one key takeaway might be that there are inklings that a subset of student-level instructional experiences subsequent to PreK may explain PreK persistence for some learning domains and may be worthy of future study in other settings. For example, we found that time spent in different types of instruction in kindergarten—primarily in the conceptual block examining constrained and unconstrained instruction across language/literacy and math—was associated with PreK persistence to some degree. Although the pattern was certainly not definitive, findings suggest that PreK persistence may be more likely when children spend a balanced

amount of time in kindergarten exposed to unconstrained and constrained instruction. The convergence pattern for literacy skills was less pronounced when children spent more time in unconstrained language instruction (controlling for time spent in constrained instruction) and less time in constrained math instruction (controlling for time spent in unconstrained instruction) in kindergarten. More time in unconstrained language instruction (controlling for constrained instruction) supported PreK persistence in language skills. The PreK boost on math skills—as measured with the REMA—was more likely to be sustained when children spent less time in constrained literacy instruction (controlling for time spent in unconstrained instruction). And this general pattern does align with constrained skill theory (Paris, 2005; Snow & Matthews, 2016), arguing that the benefits of early learning programs are more likely to persist when children experience a balance of unconstrained and constrained instruction in early childhood contexts.

Importantly, findings in support of the sustaining environments hypothesis were not always specific to outcomes matching the content delivered during this instruction (i.e., math vs. literacy/language). One interpretation of this pattern is cross-domain transfer. For example, the early PreK-induced boost in language skills was more likely to be sustained when children spent less time in constrained math instruction. Constrained math instruction in kindergarten may be rote and procedural and likely to focus on skills like number recognition, counting, arithmetic (e.g., addition, subtraction, multiplication, and division), and identification of shapes (Clements & Sarama, 2016). These activities may be less likely than more unconstrained activities (e.g., algebraic thinking, composing shapes, patterning, comparing numbers, relational thinking) to engage students in rich discussion and to facilitate verbal exchanges related to problem solving that can support critical thinking (Rittle-Johnson et al., 2015). When teachers provide more constrained math instruction, such activities may succeed in supporting those explicitly targeted skills for children who did not have an opportunity to develop them through participation in BPS PreK. Yet, cross-domain findings could also point to these individual experiences being proxies for a more general set of high-quality instructional practices. Indeed, there may be some broader latent construct of classroom quality—not captured by measures like the CLASS and teacher reports of instruction—closely associated with the extent to which teachers engage in unconstrained versus constrained instruction within and across domains (Hamre et al., 2014). In light of the large number of null interactions, however, any significant results reported here should be interpreted cautiously.

In addition, as explored in our robustness checks, an alternative argument for the study's findings could be that the instructional experiences students have in kindergarten are endogenous to their skills at kindergarten entry and actually reflect this variation in early skills. And we did find that children who arrived at kindergarten scoring higher on assessments of language were exposed to more unconstrained language instruction, less constrained literacy instruction, less constrained math instruction, more time in centers, and less time in small groups and whole groups. However, the differences were fairly minimal, falling in the range of about 5 to 10 total minutes across the full observational period of about four hours (see online supplemental materials). There were also slight differences in classroom experiences—such as emotional support and organization—favoring those with higher skills at kindergarten entry. These differences, though statistically significant, were also fairly small in magnitude. Correlations between kindergarten entry skills and kindergarten instructional experiences were all small (range from .00–.20). Taken together, we generally found that children's instructional experiences in kindergarten were more likely to be similar to one another than different and this issue of endogeneity on its own did not likely drive the main pattern of results—with the majority of the interactions being null—that we observed in these analyses.

Limitations and Directions for Future Research

Despite key strengths, this study has a number of limitations that future work should seek to address. First, as we note above, this is a descriptive study and we lack a research design that would allow for causal inference. Although we include a robust set of covariates and conduct a number of sensitivity analyses to test the robustness of our findings to potential omitted variables, there are certainly issues with selection bias in the study. Indeed, the students in the BPS PreK attender group are more advantaged than the students who did not enroll in BPS PreK, a difference that could be affecting both access to different types of learning experiences in kindergarten and the extent to which benefits of BPS PreK are sustained or not over time. Second, because we are focused on interactions in this study, we may have benefited from a larger student sample size and more statistical power. As we first noted above, we had to balance the cost of collecting individual-level observational data against the issue of statistical power and future studies may seek to allocate increased resources toward collecting data on a larger sample. Third, although our observational data are rich, we were only able to collect data on students on two separate days for about 3–4 total hours across occasions. It is possible that these days were not fully indicative of students' experiences in their kindergarten classrooms. Future work may consider a larger number of measurement occasions when assessing instruction.

Importantly, the associations that we observed in this study may vary for students from different racial/ethnic and linguistic backgrounds and students from families with low incomes. We lacked sufficient statistical power in this study to fit models that could test the associations of interest within these important subgroups. Future research with a larger and racially/ethnically and socioeconomically diverse sample of students and classrooms should be done to explore heterogeneity of the sustaining environments hypothesis for different groups of students. Finally, although we explicitly sought to conduct our study in partnership with the Boston Public Schools, it is also true that this sample and context are not necessarily generalizable to other settings and further replication work is needed.

In contrast to some earlier work in this area (e.g., Carr et al., 2019), we also did not find consistent evidence that exposure to higher levels of process quality of other classroom-level indicators of instruction sustained the benefits of PreK. It may be that in this study we were underpowered to detect interactions between BPS PreK and global observational measures like the CLASS. However, it could be that our null results for the CLASS domains are unique to the BPS setting, which is a nationally recognized example of a high-quality, scaled PreK program. For example, prior studies have found that the CLASS does not predict gains in BPS

PreK students' math, language, and executive function skills (Guerrero-Rosada et al., 2021; Weiland et al., 2013). In contrast, prior work that has identified process quality as a sustaining environment in kindergarten has generally examined lower-quality PreK programs delivered to children from families experiencing poverty in the rural United States (Carr et al., 2019) and sub-Saharan Africa (Wolf et al., 2019). Sustaining environments may not be universal, but rather specific to context, culture, and PreK and elementary school quality.

Implications

Although there is clear public and popular support for expanding access to high-quality early childhood education, there continues to be debate about why the benefits of PreK appear to fadeout/converge over time (Bailey et al., 2016). Better understanding the drivers of PreK persistence can inform interventions and supports to enhance the benefits of both PreK programs and children's elementary school learning environments. Findings from this study have implications for future research. Primarily, they highlight the importance of considering approaches beyond sustaining environments for maintaining the PreK boost. For example, targeted interventions—like those than aim to systematically align instruction across the early grades and have been shown to be effective—may be a particularly fruitful strategy for supporting PreK persistence (e.g., Mattera et al., 2018) and are worthy of continued study. In addition, future research examining effects of PreK programs across time must examine a range of student outcomes that vary in the extent to which the fadeout/convergence pattern is observed, prioritize larger studies with increased power to detect interaction effects, and consider capturing information on sustaining environments at multiple levels.

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*(12), 1103–1127. https://doi.org/10.1037/bul0000212

Ansari, A., Pianta, R. C., Whittaker, J. V., Vitiello, V. E., & Ruzek, E. A. (2020). Persistence and convergence: The end of kindergarten outcomes of pre-K graduates and their nonattending peers. *Developmental Psychology*, *56*(11), 2027–2039. https://doi.org/10.1037/dev0001115

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, *56*, 100910. https://doi.org/10.1016/j.dr.2020.100910

Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology*, *52*(9), 1457–1469. https://doi.org/10.1037/dev0000188

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

Bassok, D., Gibbs, C. R., & Latham, S. (2019). Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010? *Child Development*, *90*(6), 1875–1897. https://doi.org/10.1111/cdev.13067

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*(1), 7–34. https://doi.org/10.1086/653468

Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development*, *85*(1), 140–159. https://doi.org/10.1111/cdev.12117

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development: Handbook of child psychology* (pp. 993–1028). Wiley.

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, *12*(1), 3–9. https://doi.org/10.1111/cdep.12260

Carr, R. C., Mokrova, I. L., Vernon-Feagans, L., & Burchinal, M. R. (2019). Cumulative classroom quality during pre-kindergarten and kindergarten and children's language, literacy, and mathematics skills. *Early Childhood Research Quarterly*, *47*, 218–228. https://doi.org/10.1016/j.ecresq.2018.12.010

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., Horm, D., Jeon, S., & Ryu, D. (2019). Do stability of care and teacher–child interaction quality predict child outcomes in early head start? *Early Education and Development*, *30*(3), 337–356. https://doi.org/10.1080/10409289.2018.1546096

Claessens, A., Engel, M., & Curran, F. C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, *51*(2), 403–434. https://doi.org/10.3102/0002831213513634

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 Maths Assessment. *Educational Psychology*, *28*(4), 457–482. https://doi.org/10.1080/01443410701777272

Clements, D. H., & Sarama, J. (2016). Math, science, and technology in the early grades. *The Future of Children*, *26*(2), 75–94. https://doi.org/10.1353/foc.2016.0013

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

Connor, C. M., Adams, A., Zargar, E., Wood, T. S., Hernandez, B. E., & Vandell, D. L. (2020). Observing individual children in early childhood classrooms using Optimizing Learning Opportunities for Students (OLOS): A feasibility study. *Early Childhood Research Quarterly*, *52*(Pt B), 74–89. https://doi.org/10.1016/j.ecresq.2019.10.001

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., Piasta, S. B., Crowe, E. C., & Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, *38*(2), 85–99. https://doi.org/10.3102/0013189X09332373

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child x instruction interactions. *Journal of Educational Psychology*, *96*(4), 682–698. https://doi.org/10.1037/0022-0663.96.4.682

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology*, *98*(4), 665.

Cummings, K. D., Kaminski, R. A., Good, R. H., & O'Neil, M. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of first sound fluency. *Assessment for*

Effective Intervention, *36*(2), 94–106. https://doi.org/10.1177/1534508410392209

Duncan, S., & DeAvila, E. (1998). *PreLAS 2000 technical report*. McGraw-Hill.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Pearson Assessments.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157–178. https://doi.org/10.3102/0162373712461850

Farran, D. C., & Bilbrey, C. (2004). *Narrative record* [Unpublished instrument]. DC Farran, Peabody Research Institute, Vanderbilt University, Nashville, Tennessee.

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly*, *27*(2), 198–209. https://doi.org/10.1016/j.ecresq.2011.10.001

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2021). Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly*, *54*, 1–12.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*, *85*(3), 1257–1274. https://doi.org/10.1111/cdev.12184

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review*, *65*, 107–125. https://doi.org/10.1016/j.econedurev.2018.05.001

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E., Clements, D., Sarama, J., & Duncan, G. J. (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *Journal of Research on Educational Effectiveness*, *11*(3), 339–374. https://doi.org/10.1080/19345747.2018.1441347

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Li, W., Duncan, G. J., Magnuson, K., Schindler, H. S., Yoshikawa, H., & Leak, J. (2020). Timing in Early Childhood Education: How Cognitive and Achievement Program Impacts Vary by Starting Age, Program Duration, and Time since the End of the Program. EdWorkingPaper. *Annenberg Institute for School Reform at Brown University*, No. 20–201.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research*, *18*(2-3), 292–326. https://doi.org/10.1177/0049124189018002004

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.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*(1), 33–51. https://doi.org/10.1016/j.econedurev.2005.09.008

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*. MDRC.

McCormick, M., Hsueh, J., 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*. MDRC.

McCormick, M. P., Weiland, C., Hsueh, J., Maier, M., Hagos, R., Snow, C., Leacock, N., & 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*(Part B), 57–73. https://doi.org/10.1016/j.ecresq.2019.06.012

McCormick, M., Weiland, C., Hsueh, J., Pralica, M., Weissman, A. K., Moffett, L., Snow, C., & Sachs, J. (2021). Is skill type the key to the prek fadeout puzzle? Differential associations between enrollment in preK and constrained and unconstrained skills across kindergarten. *Child Development*, *92*(4), e599–e620. https://doi.org/10.1111/cdev.13520

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184–202. https://doi.org/10.1598/RRQ.40.2.3

Pearman, F. A., II, Springer, M., Lipsey, M., Lachowicz, M., Swain, W., & Farran, D. (2020). Teachers, schools, and pre-K effect persistence: An examination of the sustaining environment hypothesis. *Journal of Research on Educational Effectiveness*, *13*(4), 547–573. https://doi.org/10.1080/19345747.2020.1749740

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H Brookes Publishing.

Porter, K. E. (2018). Statistical power in evaluations that investigate effects on multiple outcomes: A guide for researchers. *Journal of Research on Educational Effectiveness*, *11*(2), 267–295. https://doi.org/10.1080/19345747.2017.1342887

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*(4), 269–292. https://doi.org/10.1016/j.jsp.2010.03.002

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). *Head Start Impact Study. Final report*. Administration for Children & Families.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, *2*(3), 2332858416657343. https://doi.org/10.1177/2332858416657343

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x

Sarama, J., & Clements, D. H. (2009). Concrete" computer manipulatives in mathematics education. *Child Development Perspectives*, *3*(3), 145–150. https://doi.org/10.1111/j.1750-8606.2009.00095.x

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

Schickendanz, J., & Dickinson, D. (2005). *Opening the World of Learning (OWL)*. Pearson.

Schrank, F. A., McGrew, K. S., Ruef, M. L., & Alvarado, C. G. (2005). *Batería III Woodcock-Muñoz: Assessment Service Bulletin Number 1, Overview and Technical Supplement*. Riverside Publishing.

Shapiro, A., Martin, E., Weiland, C., & Unterman, R. (2019). If you offer it, will they come? Patterns of application and enrollment behavior in a universal prekindergarten context. *AERA Open*. Advance online publication. https://doi.org/10.1177/2332858419848442

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The Future of Children*, *26*(2), 57–74. https://doi.org/10.1353/foc.2016.0012

Stipek, D., Clements, D., Coburn, C., Franke, M., & Farran, D. (2017). PK-3: What does it mean for instruction? *Social Policy Report*, *30*(2), 1–23. https://doi.org/10.1002/j.2379-3988.2017.tb00087.x

Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early grade teacher effectiveness and pre-K effect persistence: Evidence from Tennessee. *AERA Open*, *1*(4), 2332858415612751. https://doi.org/10.1177/2332858415612751

Unterman, R., & Weiland, C. (2020). *Higher-Quality elementary schools sustain the prekindergarten boost: Evidence from an exploration of variation*

in the Boston prekindergarten program's impacts (EdWorking-Paper No. 20–321). Annenberg Institute for School Reform at Brown University.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open*, *4*(1), 2332858418757735. https://doi.org/10.1177/2332858418757735

Weiland, C., Moffett, L., Guerrero-Rosada, P., Weissman, A., Zhang, K., Maier, M., Snow, C., McCormick, M. P., Hsueh, J., & Sachs, J. (2021). *Is child-level measurement the key to improving the predictive validity of observational measures in early education classrooms?* [Manuscript submitted for publication]. School of Education, University of Michigan.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199–209.

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

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

Wolf, S., Aber, J. L., Behrman, J. R., & Peele, M. (2019). Longitudinal causal impacts of preschool teacher training on Ghanaian children's school readiness: Evidence for persistence and fade-out. *Developmental Science*, *22*(5), e12878. https://doi.org/10.1111/desc.12878

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*(1), 122–154. https://doi.org/10.1002/pam.20310

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Woodcock-Johnson III diagnostic Supplement to the tests of cognitive abilities*. Riverside.

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*. Society for Research in Child Development.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, *26*(2), 21–35. https://doi.org/10.1353/foc.2016.0010

Received June 4, 2021
Revision received January 26, 2022
Accepted February 7, 2022 ■