CHILD DEVELOPMENT

Child Development, xxxx 2021, Volume 00, Number 0, Pages 1–9

The Kindergarten Hotspot: Literacy Skill Convergence Between Boston Prekindergarten Enrollees and Nonenrollees

Christina Weiland
University of Michigan

Rebecca Unterman
MDRC

Anna Shapiro
University of Michigan

Preschool improves children's kindergarten readiness, but the cognitive outcomes of preschool enrollees and nonenrollees tend to converge partially or fully in elementary school. In older programs, most of this convergence occurs in kindergarten (Li et al., 2016), but evidence from today's programs is sparse. Using data on 4,971 children who applied to the Boston Public School Prekindergarten program and a quasi-experimental approach, we examine convergence in kindergarten through third grade (K-3) literacy outcomes of prekindergarten enrollees and nonenrollees. Consistent with previous literature, most of the convergence in K-3 literacy outcomes occurred in kindergarten. Our findings suggest that detailed investigations into the kindergarten teaching and learning context may be particularly important for solving the widely noted preschool convergence pattern.

Postprogram differences in cognitive outcomes between preschool enrollees and nonenrollees shrink in elementary school, sometimes partially and sometimes fully. This convergence pattern is consistent across older and recent preschool studies in the United States, and in contexts in which the benefits of preschool attendance are detected later in the life course (Yoshikawa, Weiland, & Brooks-Gunn, 2016). Existing research shows that much of this convergence occurs very early in elementary school. For example, a recent meta-analysis found that about half of the eventual convergence on cognitive outcomes occurs during kindergarten (i.e., ages 5–6) and then by about half again by the end of second grade (Li et al., 2016).

However, most of this evidence is drawn from relatively small programs from decades ago when alternative options were different from today's and when parents of all income levels invested less time and money in their children's learning (Bassok, Finch, Lee, Reardon, & Waldfogel, 2016). Kindergarten in the United States has changed markedly as well. As documented in Bassok, Latham, and Rorem (2016), today's kindergarten teachers hold higher academic expectations for children, devote more time to advanced literacy and math content, use a more teacher-directed approach, and devote less time to other learning domains (e.g., science, music, art) compared to kindergarten teachers from just a decade earlier. The consequences of these changes are ambiguous in terms of their effects on convergence patterns. Today's home and kindergarten contexts may now be better poised to build on preschool enrollees' enhanced skills and knowledge (sustaining more of the preschool boost) or to compensate for any gaps in the skills and knowledge of nonenrollees (contributing to faster convergence).

Evidence on convergence so far from more recent large-scale programs is sparse, but in two of the largest and most rigorous recent studies, convergence was particularly rapid. In the experimental Head Start Impact Study, there were positive and

This study is funded by the Institute of Education Sciences RFA R305A140059 and R305B150012. Thanks to the Boston Public Schools (BPS), Jason Sachs, Brian Gold, the BPS Department of Early Childhood coaches and staff, the BPS Office of Data and Accountability (particularly Nicole Wagner Lam, Erin Cooley, Barry Kaufman, and Peter Sloan), Kamal Chavda, and the Wellesley Centers for Women. Special thanks also to Hirokazu Yoshikawa, Howard Bloom, Richard Murnane, David Deming, Catherine Snow, Caroline Ebanks, Gina Biancarosa, Sarah Kabay, Eleanor Martin, Shana Rochester, and Sara Staszak.

Correspondence concerning this article should be addressed to Christina Weiland, School of Education, 610 E, University Ave, Ann Arbor, MI 48104. Electronic mail may be sent to weilandc@umich.edu.

 0009-3920/2021/xxxx-xxxx
DOI: 10.1111/cdev.13499

statistically significant impacts on children's language and literacy skills at the end of the preschool year (.09 to .25 *SD*s; Puma et al., 2012). By the end of kindergarten, benefits on the same skills had declined to −.02 to .06 *SD*s and treatment-control differences were no longer statistically significant—a faster rate of decay than found in the overall literature. Likewise, the Tennessee Voluntary Prekindergarten quasi-experimental study found initial statistically significant benefits for prekindergarten enrollees (.09 to .41 *SD*s across language, literacy, and math) that did not persist through the end of kindergarten (−.10 to .09 *SD*s; Lipsey, Farran, & Hofer, 2015).

In contrast, in a nonexperimental, nationally representative study using Early Childhood Longitudinal Study-Kindergarten (ECLS-K) data, differences favoring preschool enrollees decayed at a slower rate and mirrored the pattern in the older literature for both the 1998 and 2010 cohorts in reading and math. For example, in the 2010 cohort, preschool enrollees outperformed nonenrollees by .15 *SD*s in the fall of kindergarten and by .09 *SD*s in the spring of kindergarten—a decline of approximately 40% for reading outcomes (and 54% in math; Bassok, Gibbs, & Latham, 2019). Data on kindergarten through second-grade (K-2) cognitive outcomes are not available in most studies of the longitudinal effects of preschool, including in programs with particularly strong results from third grade and beyond (e.g., Barnett, Jung, Youn, & Frede, 2013; Cascio & Schanzenbach, 2013; Gormley, Phillips, & Anderson, 2018; Ladd, Muschkin, & Dodge, 2014).

Additional evidence on convergence patterns in large-scale public preschool programs is important for understanding the effects of today's programs in a changing context, particularly at a time when children are losing ground in the United States in fourth-grade reading proficiency (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2019). Understanding *when* enrollee and nonenrollees' outcomes converge in the early elementary years may be useful for understanding how best to build on the gains from preschool and to boost student achievement overall.

We add to this nascent literature by exploring the pattern of convergence between Boston Public Schools (BPS) Prekindergarten enrollees and children who applied to the program but ultimately did not enroll in 2009–2010 and 2010–2011. In a rigorous regression discontinuity study, the program had meaningful impacts on children's language, mathematics, literacy, executive function, and socioemotional skills at kindergarten entry (the 2008–2009 cohort formed the treatment group; Weiland & Yoshikawa, 2013). In a subsequent study of children who were randomly assigned an offer to enroll in the program versus not via lotteries for oversubscribed programs in 2007–2011, there were no lasting benefits in third grade on retention, special education placement, and standardized reading and math test scores (Weiland et al., 2019). Importantly, this study sample (children who applied to oversubscribed programs) represented only 25% of all appliers and was more advantaged than the full applicant population. However, quasi-experimental research using propensity score weighting found a pattern of small persistent benefits on those same outcomes for the full sample of children who enrolled in the program between 2007 and 2011. In this study, we use K-3 literacy test scores to add to this set of findings. We explore *when* and *how rapidly* convergence of literacy outcomes occurred between kindergarten and the end of third grade in the Boston context.

Method

Setting and Program Details

The BPS Prekindergarten program began in 2005. In our study years, it was based entirely in the public schools, paid teachers on the same scale and subjected them to the same educational requirements (masters degree within 5 years) as K-12 teachers, and used proven curricula (the language and literacy-focused *Opening the World of Learning* curriculum and the mathematics *Building Blocks* curriculum). It was open to all children in the city, though demand for seats outstripped supply.

In our study's focal years in K-2, the district implemented the literacy curriculum *Reading Street* and the mathematics curriculum *TERC Investigations* which do not have as strong an evidence base as the district's prekindergarten program curricula. Observational ratings data show that prekindergarten classroom instructional quality was markedly higher on average than K-3 instructional quality (see Weiland et al., 2019).

Sample

Our analytic sample comes from the population of 4-year-old children who applied to the BPS Prekindergarten program in 2009–2010 or 2010–2011 who had at least one K-2 literacy test score (N = 4,971, or approximately 73% of the full sample

of appliers). We focus on these two cohorts because our key literacy outcome measures were collected most consistently in these years. To be conservative and in accordance with other Boston Prekindergarten studies (Weiland et al., 2019), we defined program participation as enrolling in at least one day of BPS Prekindergarten. In all, our sample consists of 3,551 enrollees and 1,420 children who applied but did not enroll (i.e., nonenrollees). Enrollee children were enrolled 169 days on average (SD = 31 days) out of a possible 180 days; only 268 enrollee children were enrolled fewer than 150 days (M = 71 days; SD = 50 days).

As shown in Table 1, the majority of our analytic sample was eligible for free-or-reduced-price lunch and either Black or Hispanic. Compared to nonenrollees, enrollee children were on average less likely to be eligible for free-or-reduced-price lunch in kindergarten (−7 percentage points, $p < .001$), less likely to be Hispanic (−13 percentage points, $p < .001$) and more likely to be Asian (4 percentage points, $p < .001$). Outcome data availability rates were very similar among enrollees and nonenrollees in our analytic sample (e.g., missing rates ranged from 8%–26% across K-2, with small differences in 0.36–7.23 percentage points by enrollment status; results available upon request).

Table 1
Background Characteristics of Boston Prekindergarten Enrollees Versus Nonenrollees

	Did not enroll	Enrolled	Difference
Age (PK year)	4.52	4.53	0.01
			(0.09)
Eligible for subdized lunch	77.03	69.90	−7.13***
			(1.31)
% Male	51.48	50.07	−1.41
			(1.59)
% Black	26.87	28.34	−1.47
			(1.30)
% Hispanic	55.60	42.23	−13.37***
			(1.48)
% Asian	4.61	8.86	4.25***
			(0.78)
% Mixed/other	2.35	2.72	0.37
			(0.51)
Home language- % Spanish	32.09	30.85	−1.24
			(1.35)
Home language- % other	19.36	21.65	2.29
			(1.25)
Country of origin- % USA	92.80	95.13	2.33**
			(0.72)
PK eligible year 2010	53.47	49.92	−3.55*
			(1.59)

Note. Means were estimated using OLS models with fixed effects for the school each student lived closest to in their first Boston Public Schools year. Statistical significance levels are indicated as: *** = .1%; ** = 1%; * = 5%. Standard errors in parentheses. N for full sample was ~4,971 (N enrollees = 3,551, N nonenrollees = 1,420).

Outcomes

K-2 Literacy Skills

We use teacher-collected data from the Dynamic Indicators of Basic Literacy Skills (DIBELS; Good & Kaminski, 2002; Good et al., 2011). Administered subtests measured children's letter knowledge (Letter Naming Fluency; LNF), oral reading fluency (Oral Reading Fluency; ORF), phonological awareness (Initial Sound Fluency and Phoneme Segmentation Fluency; ISF and PSF), and alphabetic principle (e.g., letter-sound correspondence and the ability to blend letters into words; Nonsense Word Fluency; NWF). These subtests have high reliability (0.9 or above), are widely used, are sensitive to intervention effects, and have good concurrent, predictive, and discriminant validity properties (e.g., Biancarosa, Bryk, & Dexter, 2010; Burke, Hagan-Burke, Kwok, & Parker, 2009; Good et al., 2004, 2011). Additionally, DIBELS subtests have strong concurrent and predictive validity with the primary early literacy test used in recent public preschool program evaluations, the Woodcock-Johnson Letter-Word Identification subtest (Speece, Mills, Ritchey, & Hillman, 2003).

As shown in Table 2, BPS teachers administered different subtests at different time points from K-2, as the DIBELS is designed to follow children's developmental progression in literacy. For example, ISF is a beginning reading skill measured only in kindergarten as most children will have mastered this skill by kindergarten spring. ORF, in contrast, is not measured until first grade, when it is developmentally and instructionally expected for most students.

The district switched from using the DIBELS Sixth Edition to the DIBELS Next in 2012–2013, when our second cohort was in first grade. LNF and PSF subtests that children took in first-grade fall are equatable across versions but the ORF subtests are not (Good et al., 2011). For this reason, we do not combine the first-grade end-of-year ORF scores for the two cohorts. Note that although math was also a focal domain of the prekindergarten program, data on children's K-2 math skills are not

Table 2
Descriptive Statistics for K-3 Outcomes by Enrollee Status

	M	*SD*		*M*	*SD*
Fall kindergarten			Spring kindergarten		
ISF			PSF		
Nonenrollee	11.93	9.83	Nonenrollee	36.77	16.71
Enrollee	15.46	13.25	Enrollee	39.56	15.59
LNF			NWF		
Nonenrollee	17.66	15.23	Nonenrollee	34.02	21.26
Enrollee	25.66	15.03	Enrollee	38.76	22.13
Fall first grade			Spring first grade		
LNF			ORF (Cohort 1)		
Nonenrollee	42.35	16.88	Nonenrollee	60.10	35.81
Enrollee	46.71	16.87	Enrollee	69.31	39.16
PSF			DORF (Cohort 2)		
Nonenrollee	33.27	17.41	Nonenrollee	57.34	36.12
Enrollee	35.86	15.84	Enrollee	64.81	35.91
Fall second grade			Spring second grade		
ORF			DORF		
Nonenrollee	59.00	34.38	Nonenrollee	94.75	39.12
Enrollee	67.16	35.66	Enrollee	101.63	39.08
			Spring third grade		
			ELA		
			Nonenrollee	0.08	0.99
			Enrollee	0.25	0.95

Note. Raw scores for Dynamic Indicators of Basic Literacy Skills (DIBELS) subtests and *z*-scored third-grade standardized test scores are shown. LNF = Letter Naming Fluency; ISF = Initial Sound Fluency; PSF = Phoneme Segmentation Fluency; ORF = Oral Reading Fluency (DIBELS 6th ed.); DORF = Oral Reading Fluency (DIBELS NEXT); ELA = Massachusetts Comprehensive Assessment System/Partnership for Assessment of Readiness for College and Careers English Language Arts.

available in these years (we discuss this limitation further in the Discussion section).

Third-Grade Standardized Reading Scores

Our 2009–2010 cohort took the Massachusetts Comprehensive Assessment System in third grade, the test used for state accountability purposes in Massachusetts (see Weiland et al., 2019 for psychometric details). In 2015, all but two schools in BPS chose to administer a new test based on Common Core standards, the Partnership for Assessment of Readiness for College and Careers assessment. We use students' English Language Arts scores from these tests. To pool them, we followed the state's recommendations and standardized students' estimated theta (i.e., IRT) scores (Massachusetts Department of Elementary & Secondary Education, 2016).

Covariates

Using administrative records and following other preschool studies (Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Wong, Cook, Barnett, & Jung, 2008), we constructed a set of student-level covariates. We captured students' race/ethnicity using a set of dichotomous variables that identified whether a student was Asian, Black, Hispanic, White, or mixed/other. Similarly, we used a set of dichotomous variables to identify whether the student's home language was English only, Spanish, or another language. Using student birthdates, we calculated students' age as of September 1 in their prekindergarten application year. We also created dichotomous variables that identified each student's eligibility for free-or-reduced-price lunch in prekindergarten or kindergarten; whether the student was male; and whether the student's country of origin was the United States. Finally, we identified which school they lived closest to in their first BPS year as a proxy for home neighborhood.

Analytic Approach

We used a propensity-score weighting approach to estimate the relationship between BPS prekindergarten enrollment and our key K-3 literacy

outcomes, adjusting for observable selection into enrolling in the program. Specifically, we use a logit model to predict the probability that a student would enroll in BPS prekindergarten conditional on their background characteristics, their cohort year, and home neighborhood in their first BPS year. We then inverted these propensities to obtain an inverse probability weight (IPW) that we used in our subsequent linear regression analysis to adjust for selection into the program. This approach to propensity score analysis tends to lead to less data loss than other propensity score approaches (Imbens & Wooldridge, 2009; Murnane & Willett, 2010).

The covariate differences between prekindergarten enrollees and nonenrollees shown in Table 1 were greatly reduced using IPW (e.g., the standardized raw difference on the percent of enrollee and nonenrollee students who were Hispanic and took the kindergarten fall LNF was −0.32 and the weighted difference was −0.01; results available upon request). Across outcomes and time points, no weighted covariate difference across the two groups was above .1 *SD*, the standard threshold in the field (Imbens & Rubin, 2015).

In addition to the IPW, our regression equation included, by reading outcome, a Boston Prekindergarten enrollment indicator, child covariates (gender, race/ethnicity, free-or-reduced-price lunch, home language, country of origin, cohort), and fixed effects for school they lived closest to in their first BPS year (as a proxy for home neighborhood). Note that, as explained earlier, our reading outcomes changed over time, following the developmental continuum for literacy in these years (Good et al., 2011). For analytic simplicity, we did not impute covariates in our IPW work (115 children—about 2%—were missing at least one covariate value).

Counterfactual

Information on the before-kindergarten settings of nonenrollee children (i.e., the counterfactual to which BPS Prekindergarten is being compared) was not available. However, we do have parent-reported data on the before-kindergarten settings for nonenrollee children in the two prior cohorts (i.e., 2007–2008 and 2008–2009). Although imperfect, this information helps to clarify to *what* Boston Prekindergarten is likely being compared in this study. In the 2007–2008 and 2008–2009 cohorts, 76% of prekindergarten nonenrollees enrolled in another center-based program, 6% were in family daycare, and 18% percent were at home. Of those in center-based care, 37% were in private centers, 26% in Head Start, 7% in charter schools, and 6% in other non-BPS public settings. Accordingly, Boston Prekindergarten in our study (covering the 2009–2010 and 2010–2011 cohorts) is likely being compared to a mixed counterfactual, in which the majority of children likely enrolled in other preschool programs. For details on counterfactual data in the two prior cohorts, see Weiland et al. (2019).

Results

In Table 2, we display descriptive statistics for our K-3 outcomes. Enrollees scored higher than nonenrollees at all time points. On subtests taken at multiple time points, later test scores are generally higher than earlier test scores for both groups, reflecting reading skill development over time. For example, enrollees scored 25.66 points on the LNF subtest in kindergarten fall and 46.71 points on the LNF subtest in first-grade fall. The exception is the PSF; in both groups and cohorts, children scored higher in kindergarten spring than in first-grade fall. However, this is our only spring-to-fall comparison on the same test and the lower scores in fall might be due to summer learning loss.

Figure 1 plots the estimated differences between prekindergarten enrollees and nonenrollees on our outcomes of interest grouped by the time of the assessment. We report these differences in standard deviation units. All differences were statistically significant ($p < .001$) except DORF in spring of Grades 1 and 2. In kindergarten fall, Boston Prekindergarten enrollees outscored nonenrollees on the LNF subtest by .44 *SD* and on the ISF subtest by .26 *SD*. By the end of kindergarten spring, differences were considerably smaller .17 *SD* (PSF) and .17 *SD* (NWF). Grade 1 results were similar in magnitude for two of four subtests (.17 *SD* for fall LNF and spring ORF). The two other Grade 1 subtests (fall PSF and spring DORF) and the two Grade 2 subtests (fall ORF and spring DORF) were in the 0.06–0.09 range. In third grade, the difference between Boston Prekindergarten enrollees and nonenrollees was .10 *SD* on the state standardized reading assessment.

Robustness Checks

To probe the robustness of our results, we refit our primary models: (a) using a threshold of 150 days for defining prekindergarten enrollment

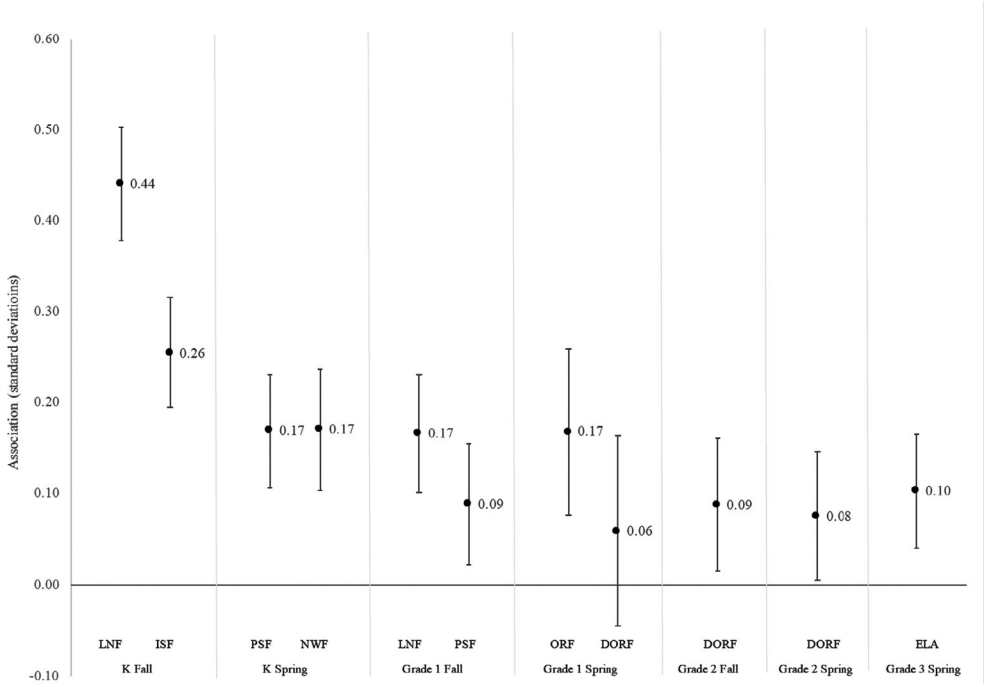

Figure 1. Associations between Boston Prekindergarten enrollment and children's K-2 literacy scores and Grade 3 standardized test scores.
Note. LNF = Letter Naming Fluency; ISF = Initial Sound Fluency; PSF = Phoneme Segmentation Fluency; ORF = Oral Reading Fluency (DIBELS 6th Edition); DORF = Oral Reading Fluency (DIBELS NEXT); ELA = Massachusetts Comprehensive Assessment System/Partnership for Assessment of Readiness for College and Careers English Language Arts; DIBELS = Dynamic Indicators of Basic Literacy Skills. Grade 1 Spring models were fit separately for Cohorts 3 and 4 because the ORF and DORF subtests are not equatable (Cohort 3 took the ORF and Cohort 4, the DORF, in Grade 1). ELA models were restricted to children who had at least one non-missing DIBELS subtest score. Estimates with corresponding confidence intervals that do not cross zero are statistically significant at the $p < .001$ level.

given the distribution in our data (e.g., 92.5% enrollees were enrolled for at least 150 days); (b) using OLS and imputing missing covariate data; (c) using fixed effects for school most attended in kindergarten; and (d) restricting our sample only to children with nonmissing outcome data on all outcome measures across K-3 (59% of the analytic sample). The pattern of results was stable across models (available upon request).

Discussion

Consistent with prior literature (Li et al., 2016), most of the test score convergence in our sample appears to have occurred during the kindergarten year. At kindergarten entry, Boston Prekindergarten enrollees scored .44 *SD* higher than nonenrollees. By the spring of kindergarten, differences had declined by about 60% (.17 *SD* for PSF and NWF). Interestingly, the second part of the broader pattern in preschool literature—that after kindergarten, the preschool advantage declines by about half again by the end of second grade (Li et al., 2016)—held for two subtests in Grade 1 and for the two Grade 2 subtests. But for two other subtests, Grade 1 magnitudes were identical to kindergarten spring magnitudes. Bassok et al. (2019) also found deviation from the broader pattern—for example, a decline of approximately half in reading and math differences between preschool enrollees and nonenrollees in the 2010 ECLS-K cohort and very little decline in these differences from end of kindergarten to end of first grade.

Overall, the convergence rates in our study and in Bassok et al. (2019) are markedly slower than those in the recent Head Start Impact Study and Tennessee Prekindergarten study (Lipsey et al., 2015; Puma et al., 2012) in which the effects present at the end of the preschool year had declined to zero by the end of kindergarten. Differences in results could be due to the different student populations. The Head Start and Tennessee programs included only children from low-income households, whereas the programs studied in ours and Bassok and colleagues' studies were not income-

limited. Although children from low-income families benefit more from public preschool programs than their peers (Yoshikawa et al., 2016), little is known about how or why convergence patterns differ by family income as older studies tended to include only children from low-income families. Another possibility is that differences in findings could be due to measurement. The Head Start and Tennessee studies relied primarily on the Woodcock-Johnson, a general achievement measure. We used the DIBELS and Bassok and colleagues used an ECLS-specific test that assessed children's basic reading skills like letter recognition, along with vocabulary and listening comprehension. It is possible that the measures we and Bassok and colleagues used may better capture the specific skills targeted in P-2 instruction.

Our results have several important limitations. Prekindergarten enrollment was not randomly assigned. Helpfully, administrative records allowed us to identify our sample as appliers to Boston Prekindergarten in the relevant years; in contrast, longitudinal nonexperimental studies of prekindergarten programs tend to lack information on applier behavior and instead all compare children enrolled in a given system K+, regardless of whether they applied to prekindergarten (e.g., Barnett et al., 2013; Jung, Barnett, Hustedt, & Francis, 2013; Phillips et al., 2017). Limiting our sample to appliers likely helped to reduce selection bias concerns to some degree. We also took steps (e.g., IPW and neighborhood fixed effects) to mitigate selection bias. However, our covariates were relatively coarse particularly compared to some other recent prekindergarten studies that have used the propensity scores approach (e.g., Gormley et al., 2018; Lipsey et al., 2015; Phillips, Anderson, & Gormley, 2018) and lacked important measures such as richness of the home literacy environment. Selection bias thus could have played a role in our results. It is worth noting that the beginning-of-kindergarten difference between enrollees and nonenrollees on the LNF test in this study (.44 *SD*) is remarkably close to estimates of the program's impact on literacy skills in the previous RD study of the program's 2008–2009 cohort (.47–.62 *SD*; Weiland & Yoshikawa, 2013). Accordingly, we view it as unlikely that our results are due entirely to selection bias. But given our study's design, we interpret our results as *associations* and not causal estimates.

In addition, 27% of appliers to Boston Prekindergarten had no DIBELS scores available, primarily because of nonenrollment in BPS in the K-2 years, and are excluded from our study. Appliers with DIBELS scores were more likely to be eligible for free-or-reduced-price lunch than appliers without (72% compared with 40%), more likely to be Hispanic (46% compared with 38%) and more likely to speak Spanish at home (31% compared with 26%). This selection into BPS limits the external validity of our results accordingly. Finally, the DIBELS does not tap the full range of reading skills children must master to become strong readers (Snow & Matthews, 2016). More work is needed on a broader range of skills, in literacy and in other domains, to inform our understanding of when and how rapidly convergence between preschool enrollees and nonenrollees occurs.

Despite these limitations, our findings have two major implications. First, in our study and more broadly, kindergarten appears to be the key period for convergence on cognitive skills for prekindergarten enrollees. This finding is consistent with the older pattern in the literature (Li et al., 2016) but is an important addition to the evidence base given that kindergarten has changed considerably over the last 15 years (Bassok, Finch, et al., 2016). Kindergarten has become more academic overall and thus is poised to either better build on gains from preschool or, alternatively, to catch up preschool nonenrollees. Our results add to findings from prior studies (Li et al., 2016) that point to the kindergarten teaching and learning context as particularly important for understanding and stemming convergence. For example, Engels, Claessens, and Finch (2013) found that kindergarten teachers tend to teach math material that the majority of students have already mastered and generally do not differentiate instruction. This may have been the case in Boston in our study years, according to a curriculum review (McCormick, Hsueh, Weiland, & Banger, 2017). More such studies are needed. And finally, though there is significant policy and practitioner attention on the prekindergarten through third-grade continuum, there are no proven P-3 models (McCormick et al., 2017; Stipek, Clements, Coburn, Franke, & Farran, 2017). Our results provide further fodder for the urgency of developing and testing P-3 models.

References

Barnett, W. S., Jung, K., Youn, M., & Frede, E. C. (2013). *Abbott Preschool Program longitudinal effects study: Fifth grade follow-up*. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from http://nieer.org/wp-content/uploads/2013/11/APPLES205th20Grade.pdf

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic gaps in early childhood experiences: 1998 to 2010. *AERA Open*, *2*, 2332858416653924. https://doi.org/10.1177/2332858416653924

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*, 1875–1897. https://doi.org/10.1111/cdev.13067

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, *2*, 2332858415616358. https://doi.org/10.1177/2332858415616358

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

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *Journal of Special Education*, *42*, 209–226. https://doi.org/10.1177/0022466907313347

Cascio, E. U., & Schanzenbach, D. W. (2013). *The impacts of expanding access to high-quality preschool education (No. w19735)*. Cambridge, MA: National Bureau of Economic Research.

Clements, D. H., Sarama, J. H., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *4*, 127–166. https://doi.org/10.5951/jresematheduc.42.2.0127

Engels, 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*, 157–178. https://doi.org/10.3102/0162373712461850

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2011). *DIBELS Next Technical Manual*. Retrieved from http://oregonrti.org/files/u11/DIBELSNext_TechnicalManual.pdf

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., . . . Flindt, N. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development (technical report, no. 7)*. Eugene, OR: University of Oregon.

Gormley, W. T., Phillips, D., & Anderson, S. (2018). The effects of Tulsa's Pre-K Program on middle school student performance. *Journal of Policy Analysis and Management*, *37*, 63–87. https://doi.org/10.1002/pam.22023

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. New York, NY: Cambridge University Press.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*, 5–86. https://doi.org/10.1257/jel.47.1.5

Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University. Retrieved from http://nieer.org/wp-content/uploads/2016/08/Arkansas20Longitudinal20Report20May2013n.pdf

Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, *33*, 162–187. https://doi.org/10.1002/pam.21734

Li, W., Leak, J., Duncan, G. J., Magnuson, K., Schindler, H., & Yoshikawa, H. (2016). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program*. Working paper.

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade (research report)*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Massachusetts Department of Elementary and Secondary Education. (2016). *Working with 2015 Massachusetts assessment data: Advisory from the Office of Planning and Research and the Office of Student Assessment Services*. Malden, MA: Massachusetts Department of Elementary and Secondary Education.

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

Murnane, R., & Willett, J. (2010). *Method matters: Improving causal inference in educational research*. New York, NY: Oxford University Press.

Phillips, D. A., Anderson, S., & Gormley, W. T. Jr (2018). The dog that didn't bark: Preschool education and middle-school attitudes in Tulsa. *Applied Developmental Science*, *22*, 200–216. https://doi.org/10.1080/10888691.2016.1263154

Phillips, D., Lipsey, M., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., . . . Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on prekindergarten effects*. Washington, DC: Brookings Institution. Retrieved from https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf

Puma, M., Bell, S. H., Cook, R., Heid, C., Broene, P., Jenkins, F., . . . Downer, J. (2012). *Third grade follow-up to the Head Start impact study final report, OPRE report # 2012-*

45. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of reading skill. *Journal of Special Education*, *36*, 223–233. https://doi.org/10.1177/00224669030360 0403

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

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2019). *National Assessments of Educational Progress (NAEP), various years, 1992–2019 Reading Assessments*. Retrieved from https://www.nationsreportcard.gov/highlights/reading/2019/

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

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

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

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