THE LONG-TERM EFFECTS OF UNIVERSAL PRESCHOOL IN BOSTON*

GUTHRIE GRAY-LOBE
PARAG A. PATHAK
CHRISTOPHER R. WALTERS

We use admissions lotteries to estimate the effects of large-scale public preschool in Boston on college-going, college preparation, standardized test scores, and behavioral outcomes. Preschool enrollment boosts college attendance as well as SAT test taking and high school graduation. Preschool also decreases high school disciplinary measures including juvenile incarceration, but has no detectable effect on state achievement test scores. An analysis of subgroups shows that effects on college enrollment, SAT-taking, and disciplinary outcomes are larger for boys than for girls. Our findings illustrate possibilities for large-scale modern, public preschool and highlight the importance of measuring long-term and non–test score outcomes in evaluating the effectiveness of education programs. *JEL Codes:* I20, J24.

I. INTRODUCTION

A substantial body of evidence establishes that early life deficits have persistent negative effects on lifetime well-being (see Knudsen et al. 2006; Almond, Currie, and Duque 2018). High-quality early-childhood interventions are seen as a promising tool to address such deficits, improve economic outcomes, and reduce socioeconomic disparities (Duncan and Magnuson 2013; Heckman 2013; Elango et al. 2016; Yoshikawa, Weiland, and Brooks-Gunn 2016; Chaudry et al. 2017). Contemporary policy efforts in the United States focus on expanding public preschool programs, many funded by state and local governments. The share of U.S. four-year-olds enrolled in state-funded preschool grew from 14% in 2002 to 34% in 2019.[1] By 2019, 44 states and 24 of the 40

*Thanks to Jason Sachs and staff at Boston Public Schools and Massachusetts Department of Elementary and Secondary Education for assistance in conducting this study. Pathak thanks the W. T. Grant Early Career Scholars Program for financial support. Eryn Heying provided superb support, and Robert Upton and Hellary Zhang provided excellent research assistance. We have benefited from comments from seminar participants at Harvard, UC Riverside, USC, UC Berkeley, UCLA, the University of Chicago, and the University of Pennsylvania.

1. For comparison, 20% of four-year-olds attended a private preschool and 7% enrolled in the federal Head Start program in 2019 (Friedman-Krauss et al. 2019).

The Quarterly Journal of Economics (2022), 1–49. https://doi.org/10.1093/qje/qjac036.
Advance Access publication on September 16, 2022.

largest U.S. cities operated large-scale public preschool programs, and nearly half of four-year-olds attended some form of publicly funded preschool (Barnett et al. 2003, Friedman-Krauss et al. 2019, NIEER 2019). Recent proposals at the federal, state, and local levels aim to continue this rapid expansion (Obama 2013; Biden 2021).[2]

Enthusiasm for public preschool derives in part from encouraging experimental evidence produced by small-scale demonstration programs in the 1960s and 1970s. The High/Scope Perry Preschool Project and Carolina Abecedarian Project randomly assigned small numbers of children to intensive preschool programs or to control groups without program access. Comparisons between the treatment and control groups show that the Perry and Abecedarian interventions improved short-term test scores and behavior as well as long-term outcomes such as educational attainment, crime, and earnings (Campbell and Ramey 1994; Schweinhart et al. 2005; Campbell et al. 2012; Heckman, Pinto, and Savelyev 2013; Garcia et al. 2020). Cost/benefit analyses suggest that these interventions are among the most cost-effective social programs on record (Barnett 1985; Belfield et al. 2006; Heckman et al. 2010b; Hendren and Sprung-Keyser 2020).

Evidence from larger-scale programs is more mixed. Nonexperimental studies of the federal Head Start program find initial test score gains that subsequently fade out, but positive effects often reappear for long-term outcomes (Currie and Thomas 1995; Garces, Thomas, and Currie 2002; Ludwig and Miller 2007; Deming 2009; Pages et al. 2020; Bailey, Timpe, and Sun 2021; Miller et al. forthcoming). This pattern may be due to persistence of effects through noncognitive channels (Heckman, Pinto, and Savelyev 2013). Two randomized trials evaluating Head Start and Tennessee's Voluntary Pre-Kindergarten Program find modest positive test score effects that fade out by elementary school (Puma et al. 2010; Puma, Bell, and Heid 2012; Lipsey, Farran, and Durkin 2018). Durkin et al. (forthcoming) report negative test score effects for the Tennessee program in sixth grade. Some analysts interpret these findings as reflecting ineffective programs, while others argue that medium-term test scores are a poor measure of program effectiveness (Mongeau 2015; Bailey et al. 2017; Heckman 2017; Whitehurst 2017). These disagreements may stem from the fact

2. President Biden's proposed American Families Plan would provide free universal preschool for all three- and four-year-olds. Ballot initiatives in Portland, St. Louis, San Antonio, and Colorado in 2020 proposed to expand public preschool.

that no study to date has used a randomized research design to study the long-term effects of a large-scale preschool program.[3]

We fill this gap by using a lottery-based research design to estimate the effects of large-scale public preschool in Boston, Massachusetts on long-term postsecondary educational outcomes. Our approach compares students who were randomly lotteried in or out of public preschool as a result of tie-breaking embedded in Boston's school assignment mechanism. We use randomized lottery offers as instruments for preschool enrollment to estimate causal effects of preschool attendance. This analysis builds on earlier work based on tie-breaking in centralized assignment systems (Abdulkadiroğlu et al. 2011, 2017) and previous studies looking at short-term impacts of preschool in Boston (Weiland and Yoshikawa 2013; Weiland et al. 2019).

We estimate causal effects of public preschool on college enrollment and persistence, grade progression and high school graduation, SAT and state achievement test scores, and behavioral outcomes related to truancy, suspension, and juvenile incarceration. Our study considers more than 4,000 randomized four-year-old applicants in seven admissions cohorts from 1997 to 2003. We measure postsecondary outcomes from a special extract of the National Student Clearinghouse, covering roughly 99% of applicants. The lottery-based research design, together with high follow-up rates for long-term outcomes covering roughly 20 years after preschool enrollment, enable us to surmount many empirical challenges with studying the long-term effects of early childhood programs.

Our analysis shows that preschool enrollment improves postsecondary outcomes. Attendance at a public preschool in Boston boosts on-time college enrollment by 8.3 percentage points, an 18% increase relative to the baseline college-going rate of 46%. Preschool enrollment leads to a 5.4 percentage point increase in the probability of ever enrolling in college and a 5.9 percentage point gain in the likelihood of ever attending a four-year college. Estimates for college graduation are also positive, though these results are less precise because some cohorts are too young to observe graduation outcomes.

To probe mechanisms for these results, we also study outcomes on the pathway to college. We find positive effects on

3. In a review of the U.S. early childhood educational literature, Cascio (2022) concludes that there is no long-term evidence on the effects of large-scale preschool programs from randomized social experiments.

several college preparatory outcomes. Preschool enrollment boosts the likelihood of graduating from high school by 6.0 percentage points. Preschool also causes an 8.5 percentage point increase in SAT test taking and raises the probability of achieving a score above the bottom quartile and in the top quartile of the SAT distribution.

We measure effects on academic achievement using scores on Massachusetts Comprehensive Assessment System (MCAS) tests and study effects on student behavior by looking at suspensions, attendance, and juvenile incarceration. We find no evidence of effects on student achievement in elementary, middle, or high school: estimated effects on MCAS scores in grades 3–10 are uniformly small and statistically insignificant. We find no impact on behavioral outcomes in middle school but significant effects on disciplinary outcomes in high school. Preschool attendance reduces the frequency of suspensions and the probability that students are incarcerated while in high school. Aggregating several measures into a summary index, we find that preschool enrollment improves high school disciplinary outcomes by 0.17 standard deviations (σ) on average.

Studies of model early-childhood demonstration programs, Head Start, and state-funded preschool programs emphasize heterogeneity in impacts by sex, race, and income (Garces, Thomas, and Currie 2002; Gormley et al. 2005; Anderson 2008; Heckman et al. 2010a; Cascio forthcoming). We examine variation in the effects of Boston's preschool program on these dimensions. The causal effects of preschool are generally larger for boys than for girls. Effects on four-year college enrollment, high school graduation, SAT-taking, and the discipline index are positive and significant for boys and insignificant for girls. Differences in estimates by race and income are generally statistically insignificant.

Our analysis makes two main contributions to the literature. First, we present the first evidence from a randomized research design on the long-term effects of a large-scale preschool program. Previous randomized studies typically look at small-scale programs (Campbell and Ramey 1994; Schweinhart et al. 2005) or are limited to short-term outcomes (Puma et al. 2010; Puma, Bell, and Heid 2012; Bitler, Hoynes, and Domina 2014; Bloom and Weiland 2015; Walters 2015; Feller et al. 2016; Kline and Walters 2016; Lipsey, Farran, and Durkin 2018; Weiland et al. 2019). Other studies look at large-scale programs using observational research designs (Garces, Thomas, and Currie 2002;

Gormley et al. 2005; Ludwig and Miller 2007; Fitzpatrick 2008; Wong et al. 2008; Deming 2009; Carneiro and Ginja 2014; Thompson 2018; Johnson and Jackson 2019; De Haan and Leuven 2020; Pages et al. 2020; Bailey, Timpe, and Sun 2021; Cascio forthcoming).[4] Studying long-term effects requires data following students over a long time horizon, which is rare among modern publicly funded preschool programs. Boston operated a large public preschool program by the late 1990s and allocated seats with a centralized mechanism, allowing us to study program impacts over multiple decades with a randomized design. The program is run by the Boston Public Schools district, so our results are relevant for evaluating expansions of preschool provided by state and local governments (Muralidharan and Niehaus 2017). Our positive estimates for educational attainment are similar to those from model demonstration programs and nonexperimental studies of Head Start, illustrating the potential for modern public preschool programs to improve long-term outcomes.

Second, our findings inform the debate regarding the link between short-term and long-term effects of education programs. Evaluations of new programs require assumptions to forecast long-term effects using short-term data (Kline and Walters 2016; Athey et al. 2019).[5] Previous evidence suggests that immediate test score gains may be a more reliable indicator of long-term effects than later test scores, and that noncognitive outcomes are an important mediator of long-term effects (Chetty et al. 2011; Heckman, Pinto, and Savelyev 2013; Chetty, Friedman, and Rockoff 2014b; Anders, Barr, and Smith forthcoming). Our results corroborate these ideas by showing positive long-term impacts for an intervention that improves adolescent behavioral outcomes but not test scores. Analyses of recent cohorts in the same Boston program studied here find initial gains during preschool on both cognitive and noncognitive assessments (Weiland and Yoshikawa 2013), but that test score effects are no longer detectable in elementary school (Weiland et al. 2019), a result that is consistent with our MCAS estimates for older cohorts. Our findings suggest

4. Related studies outside the United States include Baker, Gruber, and Milligan (2008), Havnes and Mogstad (2011, 2015), Gertler et al. (2014), Cornelissen et al. (2018), Felfe and Lalive (2018), and Baker, Gruber, and Milligan (2019).

5. Deming (2009, 112) writes: “without some sense of the connection between short-run and long-run, researchers must wait at least 15–20 years to evaluate the effect of early childhood programs.”

that such patterns may mask persistent effects on skill formation that ultimately result in higher educational attainment. More generally, our results highlight the importance of considering non–test score and long-term outcomes when assessing the effectiveness of education programs.

The rest of this article is organized as follows. The next section provides background on Boston's preschool program and describes the data used to evaluate it. Section III outlines our empirical approach and conducts validity checks on our research design. Section IV reports lottery-based estimates of preschool effects on postsecondary outcomes. Section V details results for grade progression, high school graduation, and SAT scores. Section VI reports effects on MCAS test scores and disciplinary outcomes. Section VII investigates heterogeneity across subgroups, reports effects on school and peer characteristics, and compares our estimates with results from related studies in the literature. The last section concludes. The Online Appendix provides supplementary results and additional details on data sources and sample construction.

II. Background and Data

II.A. Public Preschool in Boston

Boston Public Schools (BPS) operates separate kindergarten programs across grade levels K0 (three-year-olds), K1 (four-year-olds), and K2 (five-year-olds). Grade K2 corresponds to traditional kindergarten, and grade K0 programs enrolled a small number of students at the time of our study. We focus on K1 programs because they enroll four-year-olds, a common entry point for public preschool. Much of the growth in U.S. public preschool enrollment in recent years has also come from expansions of programs for four-year-olds (Friedman-Krauss et al. 2019).

Public preschool in Boston is universal in the sense that eligibility extends to all children residing in Boston, regardless of income.[6] As we show later, in practice the program is rationed and enrolls a relatively disadvantaged student population with high shares of nonwhite and low-income students. Programs are housed in public school facilities, including elementary schools,

6. This usage follows Cascio (forthcoming), who uses "universal" to refer to programs with no eligibility requirements beyond age.

early learning centers, and special school facilities covering early grades. BPS preschools are staffed by teachers who hold either bachelor's or master's degrees and must complete the same certification requirements as BPS teachers in higher grades. In a survey of 43 randomly selected K1 classrooms during the 2005–2006 school year, Marshall, Roberts, and Mills (2006) found that all K1 teachers held bachelor's degrees and 56% held master's degrees. In contrast, around one-third of Head Start teachers had bachelor's degrees in 2001 (Kline and Walters 2016). On average, K1 teachers had eight years of experience in BPS and six years at their current school. More than half of the teachers were non-Hispanic white, 11% were Hispanic or Latino, 10% were Black, and 6% described themselves as biracial.

During the time period of our study (1997–2003), BPS operated under an "autonomous district model," giving school principals freedom to hire teachers and choose curricula. Many programs used the Harcourt Trophies curriculum and later switched to Opening World of Learning (OWL) and Building Blocks (Schickendanz and Dickinson 2005; Clements and Sarama 2007).[7] Class sizes ranged from 10 to 25 students, with an average of 19 (Marshall, Roberts, and Mills 2006). BPS Children First estimates that the program costs roughly $13,000 (2020 dollars) for full-day preschool and about half as much for half-day programs (Massachusetts Department of Elementary and Secondary Education 1995). For comparison, state-funded preschool programs typically cost about $6,000 per student, and the federal Head Start program costs roughly $11,000 per enrolled child (Friedman-Krauss et al. 2019; U.S. DHHS 2019).

Our study period coincides with substantial changes in Boston's preschool program. In 1997, all K1 programs were half day, with students attending preschool in either the morning or the afternoon for two and a half hours. In 1997, the Boston school committee decided to partially phase out half-day K1 programs in favor of offering full-day, six-hour kindergarten for all five-year-olds (BPS 1997, 1998). As a result, the number of K1 seats declined from roughly 2,500 to 900 and the number of programs dropped from about 60 to 45 between 1997 and 1998 (Figure I). In

7. Marshall, Roberts, and Mills (2006) report that in 2005, 60% of K1 classrooms used OWL, 40% used Building Blocks, 20% used Harcourt Trophies/Reading First, 20% used a self-developed curriculum based on best practices in the field, 12% used TERC Investigations, and 8% used Readers and Writers Workshop.

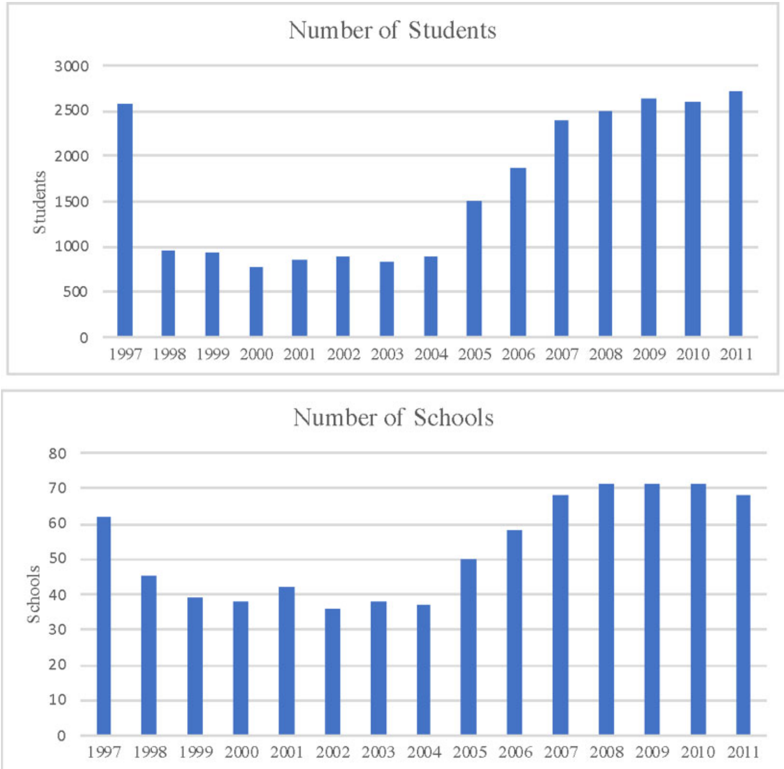

FIGURE I

Boston Preschool Students and Schools by Year (Four-Year-Olds)

This figure plots the number of four-year-old students enrolled in Boston public preschools (top panel) and the number of schools offering preschool for four-year-olds (bottom panel) by year.

1998, Boston opened three new Early Education Centers offering full-day programs. The district also opened five additional full-day programs over the next few years, resulting in a mix of full-day and half-day K1 programs.

Changes to the program continued after our study period. In 2005, Boston mayor Thomas Menino proposed expanding the supply of K1 seats citywide, and Figure I shows that Boston preschool grew to about 2,500 K1 students a year by 2008. Duncan and Murnane (2014) recount that the expansion created problems with staffing, stating: “As is common in districts that dramatically increase the supply of preschool education, BPS had difficulty finding enough suitable classrooms and trained teachers." BPS commissioned a survey in the 2005–6 school year, which documented issues with instruction, sanitation, and safety (Marshall, Roberts, and Mills 2006).[8] Around that time, BPS created a new Department of Early Childhood with new leadership, which emphasized standardized procedures and curricula across schools (Sachs and Weiland 2010). Weiland and Yoshikawa (2013) describe more details of the curriculum and program implementation in Boston, focusing on this period.

The changes that occurred after the rapid expansion of preschool in Boston are of unclear relevance to our study period. Prior to these changes, the preschool program was heterogeneous and not subject to standardized oversight. Initial surveys on program quality focus on the period right after the large expansion in 2005. Marshall and Roberts (2010) describe sustained improvements occurring between three surveys conducted in 2006, 2008, and 2010, involving space and furnishings, program structure, emotional and social supports, and instructional supports. Duncan and Murnane (2014) describe a quality-improvement strategy focusing on educational supports, staffing, coaching, and professional development. The possibility of improvement along these dimensions suggests the preschool program was lower quality according to traditional metrics during our study period.

At the same time, public programs implemented at scale elsewhere may be more likely to share features with the evolving BPS program during our study period than the carefully designed and centralized program implemented post-2005, which Duncan and Murnane (2014) describe as a "top-of-the-line model." For example, Gupta et al. (2021) identifies four threats to scaling and generalizability of early childhood programs: (i) inferential issues due to small-scale experimental studies, (ii) nonrepresentative populations, (iii) nonrepresentativeness of policy implementation, and (iv) spillover effects. As we describe in further detail below, our sample of randomized applicants is large and has similar attributes to students enrolled in the program. Potential unevenness or even mediocre implementation may be more likely in a large urban city context than the well-controlled implementations expected in model or demonstration programs. Furthermore, the time period of our study predates the large-

8. The results of this survey attracted attention on the front page of the *Boston Globe* (Jan 2007).

scale expansion of Boston's charter sector, which could be a confounding fallback option during more recent time periods.[9] Therefore, on these dimensions, our study may provide a more accurate measure of potential long-term effects on educational attainment of scaled-up modern-era preschool programs. Our discussion of how our results compare to other preschool evaluations also show remarkable consistency of the effects of preschool programs from different settings.

In recent years BPS preschools score highly on observed metrics of program quality, receiving 8 out of 10 benchmarks and ranking sixth out of 40 city-wide programs in a recent NIEER report (NIEER 2019). Recent administrations have attempted to provide enough capacity for all of Boston's four-year-olds, but as of 2019 there was only enough capacity to serve roughly half of Boston's four-year-old students (Martin 2021). The rationing of BPS preschool seats is a key element of our research design.

II.B. Data and Sample

The BPS district provided data covering all preschool applicants from fall 1997 to fall 2003. The application files contain demographics, address information, and the inputs used to implement the school assignment algorithm (described further below), including students' rank-ordered choices over schools, admission priorities, and random tie-breaking numbers. BPS also provided a second postapplication file recording school assignments, BPS preschool enrollment, and applicant names and dates of birth, which we link to the application file using a unique BPS identifier.

We measure outcomes for BPS preschool applicants by matching the applicant records to several additional data sources. Our key outcomes are measures of college attendance, college type, and college graduation derived from a special National Student Clearinghouse (NSC) data extract. We submitted names and dates of birth for BPS applicants for matching to the NSC in spring 2020. Dynarski, Hemelt, and Hyman (2015) reports that the NSC covered more than 90% of U.S. undergraduate institutions as of 2011, the earliest year of college enrollment for our cohorts, and 95% of Massachusetts undergraduate institutions.

9. Setren (forthcoming) describes that only 4 out of 10 of Boston's elementary charter schools were open before 2010.

We use the NSC records to construct two sets of postsecondary outcomes distinguished by the timing of measurement. The "on time" concept refers to whether a student achieves an outcome within a data window that assumes normal academic progress from their initial application. For example, a student who applied to preschool in fall 1997 would finish 12th grade on time in spring 2011, enroll in college on time by fall 2011, and graduate from a four-year college on time by summer 2015. The "ever" concept records the same outcomes with no restrictions on the follow-up window, which allows us to capture late enrollment but implies a shorter data window for more recent applicant cohorts. Online Appendix Table A1 summarizes the data windows available to measure outcomes for each cohort. The primary outcomes for our study are on-time enrollment in any college and whether a student ever enrolled in any college. We also report effects on type of college attended (two-year or four-year, private or public, and Massachusetts college), the total number of semesters, and college graduation. Since students applying for preschool in 2002 and 2003 would not finish a four-year college on time until after our NSC search, we do not observe college graduation outcomes for these cohorts.

Outcomes prior to college enrollment are measured by linking preschool applicants to administrative data from the Massachusetts Department of Elementary and Secondary Education (DESE). This database contains school enrollment records, demographics, and MCAS test scores in grades 3–8 and 10 for students enrolled at Massachusetts public schools. Our primary test score analysis stacks all observed test scores across grades, though we also report estimates separately by grade. The DESE data also record disciplinary outcomes including suspensions, truancy, and codes for students in juvenile incarceration, as well as SAT scores and high school graduation for Massachusetts public high schools. The primary outcome for our analysis of disciplinary outcomes is a summary index formed as the first principal component of this list of outcomes. The Online Appendix provides further information on the procedures used to clean and link data sets and construct outcomes.

Table I summarizes the characteristics of our sample of BPS preschool applicants. Column (1) displays statistics for the sample of 8,786 first-time BPS preschool applicants who applied for a K1 slot between 1997 and 2003. As shown in Panel A, nearly three-quarters of preschool applicants are Black or Hispanic, 11%

TABLE I

DESCRIPTIVE STATISTICS AND COVARIATE BALANCE

	Average characteristics		Offer differentials	
	All applicants (1)	Randomized applicants (2)	No controls (3)	Risk controls (4)
Panel A: Applicant demographics				
Black	0.432	0.407	– 0.011	– 0.015
			(0.011)	(0.017)
White	0.166	0.149	– 0.012	– 0.023*
			(0.008)	(0.012)
Hispanic	0.291	0.344	0.036***	0.020
			(0.011)	(0.015)
Female	0.495	0.488	0.011	0.060***
			(0.011)	(0.020)
Age at enrollment	4.569	4.580	– 0.025	– 0.031
			(0.017)	(0.031)
Bilingual Spanish	0.108	0.187	0.044***	0.004
			(0.008)	(0.005)
Panel B: Application characteristics				
Number of programs ranked	3.055	2.949	– 0.098***	0.041
			(0.028)	(0.038)
Walk zone	0.215	0.176	0.154***	– 0.005
			(0.010)	(0.005)
Panel C: Neighborhood characteristics				
Population	1,269.0	1,263.4	– 5.8	50.4**
			(12.7)	(22)
Median family income	39,860.8	39,907.7	– 995.6**	493.6
			(388.9)	(637.8)
Poverty rate	0.215	0.212	0.011***	0.000
			(0.003)	(0.005)
Share Black	0.374	0.384	0.033***	– 0.007
			(0.007)	(0.009)
Share white	0.382	0.370	– 0.036***	0.011
			(0.007)	(0.009)
Share Hispanic	0.210	0.220	– 0.007**	– 0.005
			(0.003)	(0.005)
Sample size	8,786	4,215	8,786	4,215

Notes. This table displays average characteristics and differences in characteristics by offer status for applicants to BPS K1 programs from 1997 to 2003. Panel A shows results for applicant demographics, Panel B reports on application characteristics, and Panel C shows results for characteristics of an applicant's block group measured in the 2000 U.S. census. Column (1) shows characteristics for all applicants, and column (2) shows characteristics for applicants subject to random assignment (those with assignment propensity scores strictly between zero and one). Column (3) reports coefficients from regressions of each characteristic on an offer indicator, controlling for year indicators. Column (4) adds controls for assignment risk and restricts the sample to applicants subject to random assignment. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

are classified as bilingual Spanish, and the typical applicant is 4.6 years old at the time of potential preschool enrollment. On average, students rank three schools on their application forms (Panel B, column (1)). Panel C displays neighborhood characteristics measured by matching an applicant's geographic information to block groups in the 2000 U.S. census. The average applicant lives in a neighborhood with a median family income of $40,000 and a poverty rate of 22%.

As discussed in the next section, our analysis focuses on applicants subject to random assignment, defined as those whose preschool offers are determined by a random tie-breaker. Table I, column (2) shows descriptive statistics for the randomized subsample. In total, 4,215 applicants are subject to random assignment. Characteristics of randomized applicants are generally similar to those of the full applicant population.

III. Empirical Framework

III.A. Research Design

Our research design relies on random tie-breaking within Boston's centralized school assignment mechanism. Households applying to BPS preschools submit rank-ordered lists of preferences for preschool programs to the district. Applicants receive priorities at each program based on sibling status and geographic proximity (those within a program's "walk zone" receive higher priority). In priority groups, tie-breaking is based on a random number assigned by the district. The mechanism combines preferences, priorities, and random tie-breakers to output a single assignment for each applicant, which is either a specific BPS preschool program or no program. During our study period, the city used the immediate acceptance (or "Boston") mechanism to determine assignments (Abdulkadiroğlu and Sönmez 2003).

Differences in assignments between students with the same preferences and priorities arise solely because of the random tie-breaking number. Few students share all the same preferences and priorities, but in practice the probability of an offer depends on a coarser set of school-level cutoffs. This motivates a strategy of controlling for the assignment propensity score, defined as the conditional probability of a preschool offer given an applicant's preferences and priorities (Abdulkadiroğlu et al. 2017). The propensity score theorem of Rosenbaum and Rubin (1983) implies that if preschool offers are random (independent of potential outcomes) conditional on preferences and priorities, then offers are also random conditional on the assignment propensity score.

A special feature of the centralized school assignment setting is that the propensity score can be calculated with knowledge of preferences, priorities, and the structure of the assignment algorithm. We compute the assignment propensity score using an analytic large-market approximation derived by Abdulkadiroğlu et al. (2017), which yields closed-form solutions for lottery number cutoffs determining whether each student is assigned to each program.[10] A student's propensity score for a particular program is the likelihood that their lottery number falls between the relevant cutoffs.[11] Our data include the random tie-breaking numbers used in the actual assignments, so we similarly code preschool offers based on whether a student's realized random number fell in the relevant cutoff region.[12]

We use preschool assignment as an instrument for preschool enrollment, controlling for the assignment propensity score. The primary estimating equations for our analysis

10. Abdulkadiroğlu et al. (2017) derive the propensity score for the deferred acceptance (DA) algorithm. Appendix A.10 of their paper shows that it is possible to construct the propensity score for the immediate acceptance algorithm by redefining priorities so that priority groups at a school consist of applicants who share original priority status at the school and rank it in the same position, then applying the formula for the DA propensity score.

11. We compute the probability of assignment to any preschool by summing the propensity score associated with an offer at each ranked preschool program. This method allows us to isolate all randomly generated offers from the assignment mechanism and extract a greater number of applicants subject to random assignment than approaches that only consider first choices (see Abdulkadiroğlu et al. 2011; Weiland et al. 2019). A histogram of the propensity score appears in Online Appendix Figure A1. Among applicants with propensity scores strictly between zero and one, 3,791 of 4,215 students (90%) have propensity score values with both offered and nonoffered applicants, implying substantial common support in this sample.

12. Online Appendix Table A2 shows that this coding replicates 94% of observed assignments. In 1997–1999, BPS used racial rebalancing to modify a small number of assignments after running the assignment algorithm, a practice that aimed to reduce segregation in Boston (Willie and Alves 1996). These postassignment moves drive the lower replication rates in 1997–1999 but do not contaminate our research design because our coding disregards rebalanced offers. Any differences between our coding of offers and final student assignments can be interpreted as noncompliance with the assignment algorithm.

are:

$$Y_i = \beta D_i + \sum_p \alpha_p 1\{P_i = p\} + X_i'\gamma + \epsilon_i, \tag{1}$$

$$D_i = \pi Z_i + \sum_p \delta_p 1\{P_i = p\} + X_i'\psi + \eta_i, \tag{2}$$

where Y_i is an outcome for student i, D_i indicates BPS preschool attendance, and Z_i indicates an offer to any BPS preschool program. Both equations include a saturated set of indicators for values of the propensity score P_i, which measures the probability of an offer to any Boston preschool (computed by summing the propensity scores for each program). We refer to the propensity score as a "risk" control. Baseline covariates X_i include race, sex, and bilingual Spanish indicators. The parameter of interest is β, which represents the causal effect of preschool attendance.

We estimate equations (1) and (2) by two-stage least squares (2SLS) in the sample of randomized applicants (those with values of P_i strictly between zero and one). The first stage fits equation (2) by ordinary least squares (OLS) and constructs predicted values $\hat{D}_i$. The second stage fits equation (1) by OLS after substituting $\hat{D}_i$ for D_i. The resulting 2SLS coefficient can be interpreted as a weighted average of local average treatment effects (LATEs) for "compliers" who enroll in BPS preschool in response to offers at each value of the propensity score (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996). Specifically, we have

$$\beta = \sum_p \left(\frac{f_p \pi_p p(1-p)}{\sum_{p'} f_{p'} \pi_{p'} p'(1-p')} \right) \beta_p, \tag{3}$$

where f_p is the fraction of students with propensity score $P_i = p$, π_p is the first-stage coefficient from a regression of D_i on Z_i at this value of the propensity score, and β_p is the coefficient from an instrumental variables (IV) regression instrumenting D_i with Z_i among students with $P_i = p$.[13] Under the standard IV assumptions of Imbens and Angrist (1994), β_p equals the LATE for compliers

13. See Appendix C of Walters (2018) for a derivation of this weighting formula. As noted by Abdulkadiroğlu et al. (2017), this estimand coincides with the coefficient from a model controlling only for an additive linear term in P_i. To see this, note that by the Frisch-Waugh-Lovell Theorem 2SLS is equivalent to

with propensity score p. As long as a preschool offer weakly increases preschool attendance at each value of the propensity score ($\pi_p \geqslant 0$), the weights in equation (3) are all positive and β gives a convex weighted average of LATEs.[14] It's worth noting that since few applicants are able to attend BPS preschool without an offer, the LATE in this setting essentially coincides with the effect of treatment on the treated (TOT) at each value of P_i (Bloom 1984).[15]

III.B. Counterfactual Preschool Choices

Kline and Walters (2016) show that the LATE associated with public preschool attendance captures a treatment effect relative to a mix of counterfactual alternatives for compliers, which may include other preschools. The interpretation of the effects we estimate therefore depends on the mix of counterfactual preschool options for compliers in our study. Because we do not directly observe counterfactuals for compliers, we turn to aggregate administrative data from BPS, Head Start enrollment data from Head Start program information reports, private school enrollment data from the NCES Private School Survey (PSS), and Census/ACS data on Boston's four-year-old population to describe preschool options during our time period. The data appendix section of the

bivariate IV using the residual from a regression of the excluded instrument on the included controls as the instrument. Since $E[Z_i|P_i, X_i] = P_i$, the residuals generated by controlling for linear and saturated propensity scores coincide and equal $Z_i - P_i$, with or without controls for other baseline covariates X_i. This logic also implies that the recentering approach of Borusyak and Hull (2021), which uses $Z_i - P_i$ as an instrument, produces the same estimand as well. Therefore, any differences between saturated, linear, and recentered 2SLS are due to finite sample issues. Online Appendix Table A8 shows that in practice we obtain very similar results with a linear control for P_i.

14. This reflects the result of Blandhol et al. (2022) that a LATE interpretation of 2SLS generally requires the first-stage specification to be "monotonicity-correct," meaning that the regression specification correctly describes the sign of the behavioral response to the instrument for each value of the covariates. In the BPS preschool setting it seems plausible that a preschool offer weakly increases the likelihood of preschool attendance for all subgroups.

15. Online Appendix Table A10 investigates effect heterogeneity by interacting the propensity score with the pre-K treatment indicator, adding the interaction between the offer and the propensity score as a second instrument. For key outcomes, we find statistically insignificant interactions with the propensity score, suggesting that the weighting scheme used to aggregate effects across values of the score is unlikely to matter. Abdulkadiroğlu et al. (2017) explore alternative weighting schemes in their study of Denver's charter schools and arrive at a similar conclusion.

Online Appendix provides more details on processing of these files. Online Appendix Table A3 reports enrollment counts in BPS preschools, private preschools, and Head Start centers, along with the population of Boston four-year-olds, for the period 1997–2019. This descriptive exercise reveals that a large majority of Boston students enrolled in some type of preschool program during our sample period, suggesting that private and Head Start centers are likely to be an important part of the counterfactual for our analysis.

Next we leverage the expansion of the BPS preschool program over time to get a better sense of the relevant counterfactual for the BPS preschool population. Online Appendix Table A4 reports coefficients from regressions of annual Head Start and private preschool enrollment shares on the share of students enrolled in BPS preschools. We find that expansions of BPS preschools coincide with significant declines in both Head Start and private preschool enrollment, with somewhat larger declines in the Head Start share. Under the strong assumption that changes in BPS enrollment are unrelated to changes in other determinants of these shares, our estimates suggest that 33% of new BPS preschool students are drawn from Head Start and 29% are drawn from private preschools, implying that the remaining 38% are drawn from no preschool. This exercise uses a different source of variation than our randomized research design, but it again suggests that other preschools are likely to serve as an important part of the counterfactual for our study.[16] As Kline and Walters (2016) show, LATE represents the policy-relevant parameter for evaluating expansions of BPS preschool regardless of the mix of counterfactuals when alternative programs are not rationed. Our estimates are therefore relevant to policy debates regarding the expansion of public preschool in Boston, at least during our study period.

III.C. Balance and Attrition

Before presenting the main estimates, we turn to two tests of the validity of our research design. Table I checks whether predetermined characteristics are balanced between offered and nonoffered students, as would be expected under random

16. Weiland et al. (2019) report that a majority of students lotteried out of Boston preschools in more recent years attend other center-based preschool programs (most commonly private preschools).

assignment. Column (3) reports coefficients from OLS regressions of student characteristics on an offer indicator, controlling for cohort indicators but not adjusting for assignment risk. These contrasts show significant imbalances by Hispanic status, bilingual Spanish, and application and neighborhood characteristics, likely because students from different demographic groups and neighborhoods apply to different programs. Column (4) restricts the sample to randomized applicants and adds controls for the assignment propensity score. Controlling for risk eliminates most of the statistically significant imbalances from column (3), illustrating the balancing properties of the assignment propensity score. We see some imbalance in offers by gender and race when controlling for risk, which we address by controlling for these variables in the analysis to follow.

Next we investigate follow-up rates for our key outcomes. Even with random assignment of preschool slots, nonrandom attrition may compromise the comparability of lottery winners and losers, possibly generating selection bias. This has been a major concern in studies of preschool programs, where long time intervals between interventions and outcomes create the potential for substantial attrition (see Armor 2014; Elango et al. 2016). This possibility motivated us to conduct a custom search of NSC records for all preschool applicants based on the names and dates of birth provided by BPS.

Table II, column (1) shows that information for roughly 99% of control group (nonoffered) applicants was submitted to the NSC. This establishes that overall attrition for postsecondary outcomes in our study is very low. As shown in Table II, column (2), applicants who were offered a preschool seat were 0.8 percentage points more likely to be submitted to the NSC. This reflects a slight imbalance in the availability of names and dates of birth in the (post-treatment) data file we received from BPS, likely because missing information was updated for a few applicants who enrolled in preschool. With 4,215 randomized applicants and an offer rate of about one-third, the 0.8 percentage point gap corresponds to 11 extra offered students. Our impact estimates for postsecondary outcomes are unlikely to be affected by this small difference in follow-up.

To measure earlier outcomes available during time enrolled in Massachusetts public schools, we link preschool applicants based on name and date of birth to records from the state's administrative database, known as the Student Information Management

TABLE II
ATTRITION

	Nonoffered follow-up rate (1)	Offer differential (2)
Name submitted to NSC	0.987	0.008**
		(0.003)
		4,215
Ever observed in SIMS	0.910	0.028***
		(0.010)
		4,215
Any MCAS score	0.845	0.038***
		(0.013)
		4,215
Number of MCAS scores	9.052	0.520***
		(0.184)
		4,215

Notes. This table reports follow-up rates and offered/nonoffered differences for key outcomes. The sample includes all randomized BPS preschool applicants. Column (1) displays the fraction of nonoffered applicants observed in each sample. Column (2) reports coefficients from regressions of follow-up on an offer indicator with controls for assignment risk. * significant at 10%; ** significant at 5%; *** significant at 1%.

System (SIMS). About 91% of nonoffered applicants are observed in the SIMS file, and applicants who receive offers are 2.8% more likely to be observed. This difference may reflect a causal effect of BPS preschool on the likelihood of attending a Massachusetts public school, perhaps because public preschool enrollment increases attachment to the public education system. This result is consistent with other evidence that attending public preschool increases the likelihood of subsequently attending a public school (Weiland et al. 2019; Greenberg et al. 2020). Similarly, we are 3.8 percentage points more likely to observe a follow-up test score for students assigned to preschool, and we see an average of 0.5 more scores for the treatment group (out of a total of up to 14 math and reading scores in grades 3–8 and 10). As a result of these modest but statistically significant differences in attrition, results for test score and behavioral outcomes derived from the Massachusetts administrative data should be interpreted with more caution than results for our primary postsecondary outcomes.[17]

17. Online Appendix Table A5 investigates balance of observed characteristics in postattrition samples matched to the state data, including covariates that are not observed for unmatched students (free and reduced-price lunch status, special education, and limited English proficiency). Characteristics are generally similar for offered and nonoffered students but we see a few imbalances in the state

IV. Effects on Postsecondary Outcomes

Boston preschool attendance increases on-time college enrollment. We arrive at this result in Table III, which reports 2SLS estimates of equations (1) and (2) for postsecondary outcomes. Column (2) reports estimates of the first-stage coefficient π, which show that a preschool offer increases the probability of preschool attendance by 65 percentage points. Column (3) displays the reduced-form effect of an offer on the outcome, estimated by replacing D_i with Y_i on the left-hand side of equation (2). A preschool offer increases on-time college enrollment by 5.4 percentage points. Since the 2SLS model is just identified, the 2SLS estimate in column (4) equals the ratio of the reduced form to the first stage, which reveals that enrollment at a Boston preschool increases on-time college enrollment by 8.3 percentage points. This estimate, which is statistically significant at the 1% level, implies an 18% increase in on-time college enrollment relative to the 46% rate for nonoffered students (column (1)).

The second and third rows of Table III, column (4) show that preschool enrollment increases on-time enrollment in four-year colleges by 8.6 percentage points ($p < .01$), but has no effect on on-time enrollment in a two-year college. Preschool enrollment increases the likelihood of on-time enrollment at a Massachusetts college by a highly significant 8.7 percentage points ($p < .01$).[18] We find positive point estimates for both public and private colleges, though only the private college estimate is marginally statistically significant.

The positive on-time enrollment effects of Boston preschools translate into positive effects on ever attending college. Table III, columns (5)–(8) display results for postsecondary outcomes measured at any time (the "ever" outcome concept). The nonoffered college attendance rate increases from 46% to 65% when we drop the on-time restriction, implying that many students enroll late. The 2SLS estimates show a slight convergence between treatment and control groups when we include late enrollment: the

outcome samples, such as a higher rate of free or reduced-price lunch eligibility among offered students.

18. The most common colleges attended by BPS preschool students in our sample are Bunker Hill Community College; the University of Massachusetts campuses at Amherst, Boston, and Dartmouth; Newbury College; Framingham State; Cambridge College; Ben Franklin Institute of Technology; and the Urban College of Boston.

TABLE III

EFFECTS OF PRESCHOOL ATTENDANCE ON POSTSECONDARY OUTCOMES

	Enrollment on time				Enrollment at any time			
	Nonoffered mean (1)	First stage (2)	Reduced form (3)	2SLS estimate (4)	Nonoffered mean (5)	First stage (6)	Reduced form (7)	2SLS estimate (8)
Any college	0.459	0.645***	0.054***	0.083***	0.650	0.645***	0.035*	0.054*
		(0.015)	(0.019)	(0.030)		(0.015)	(0.019)	(0.029)
	2,669		4,175		2,669		4,175	
Two-year college	0.096		−0.002	−0.003	0.291		0.019	0.030
			(0.011)	(0.017)			(0.018)	(0.028)
	2,669		4,175		2,669		4,175	
Four-year college	0.367		0.056***	0.086***	0.506		0.038*	0.059*
			(0.019)	(0.030)			(0.020)	(0.030)
	2,669		4,175		2,669		4,175	
Massachusetts college	0.328		0.056***	0.087***	0.504		0.045**	0.071**
			(0.019)	(0.029)			(0.020)	(0.030)
	2,669		4,175		2,669		4,175	
Public college	0.257		0.017	0.027	0.474		0.033*	0.051*
			(0.018)	(0.027)			(0.020)	(0.031)
	2,669		4,175		2,669		4,175	

TABLE III
CONTINUED

	Enrollment on time				Enrollment at any time			
	Nonoffered mean (1)	First stage (2)	Reduced form (3)	2SLS estimate (4)	Nonoffered mean (5)	First stage (6)	Reduced form (7)	2SLS estimate (8)
Private college	0.203		0.036**	0.056**	0.316		0.015	0.024
			(0.016)	(0.025)			(0.018)	(0.028)
	2,669		4,175		2,669		4,175	
Number of semesters					5.954		0.367*	0.569*
							(0.209)	(0.322)
					2,669		4,175	
Graduation	0.210	0.621***	0.010	0.016	0.325	0.621***	0.033	0.052
		(0.017)	(0.019)	(0.030)		(0.017)	(0.021)	(0.034)
	2,108		3,281		2,108		3,281	
Graduation from four-year	0.207		0.005	0.008	0.297		0.022	0.035
			(0.018)	(0.029)			(0.020)	(0.033)
	2,108		3,281		2,108		3,281	

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on postsecondary outcomes. Outcomes in columns (1)–(4) measure college enrollment within six months of a student's projected high school graduation date given their BPS preschool application year. Columns (5)–(8) display results based on college attendance at any time. On-time graduation equals one if a student graduates from a four-year college by the end of the fourth calendar year after projected high school graduation or a two-year college by the end of the second calendar year after projected high school graduation. The sample for graduation excludes students who applied to preschool in 2002 or 2003, who would not finish a four-year college on time given normal academic progress. Columns (1) and (5) show mean outcomes for nonoffered students. Columns (2) and (6) display coefficients from regressions of preschool attendance on the preschool offer. Columns (3) and (7) show coefficients from regressions of outcomes on the offer. Columns (4) and (8) report 2SLS coefficients instrumenting preschool attendance with the offer. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

estimated effect on ever enrolling at any college equals 5.4 percentage points ($p < .1$). For enrollment at any time we see large estimated effects on four-year enrollment (5.9 percentage points, $p < .1$) and enrollment at Massachusetts institutions (7.1 percentage points, $p < .05$). Estimates for both public and private institutions are positive, but the effect for public institutions is larger when we include late enrollment (5.1 percentage points, $p < .1$). Adding up the number of semesters enrolled across all institution types, we find that BPS preschool attendance increases postsecondary enrollment by 0.57 semesters, a 10% gain relative to the control mean of 6 semesters ($p < .1$).

The bottom rows of Table III display estimates for college graduation. Because we do not observe graduation outcomes for two out of seven applicant cohorts, the sample size for these outcomes falls by roughly 20%. We see no impact on the likelihood of graduating college on time (column (4)). The estimated effect on ever graduating from any college in column (8) suggests that Boston preschool enrollment increases graduation by 5.2 percentage points, a quantitatively large estimate equal to 16% of the control graduation rate of 33%. The estimate for four-year college graduation is also large at 3.5 percentage points (12% of the control mean). However, we cannot reject that these graduation effects equal zero due to a lack of statistical precision.

Our results for postsecondary outcomes are robust to several reasonable changes in estimation strategy. The imbalance by sex shown in Table I motivates a sensitivity analysis that drops the control for female from equations (1) and (2). Online Appendix Table A6 shows that dropping this control leads to slightly larger estimates than those in Table III but does not change the key results. Similarly, Online Appendix Table A7 shows that our results are robust to dropping the 1997 applicant cohort, which attended BPS preschool before the restructuring discussed in Section II.A. Online Appendix Table A8 demonstrates that results are very similar when we control linearly for values of the assignment propensity score rather than a saturated set of indicators. Finally, Online Appendix Table A9 reports an analysis that replaces the offer indicator Z_i with the randomly assigned tie-breaking number as the instrument in equation (2). Using the random number as the instrument reduces statistical power because this simpler strategy does not fully exploit the structure of the assignment mechanism, but the pattern of estimates and statistical significance is similar to our baseline results in Table III. Estimated effects on college graduation at any time are marginally statistically significant in specifications without the female control, with linear controls for the values of the assignment propensity score, and using the random number as the instrument.

Taken together, our results reveal a clear pattern of positive effects of preschool attendance on postsecondary educational outcomes for students in Boston. These findings are noteworthy in light of large gaps in graduation rates and time to degree by race and income (Bowen and Bok 2000). Our results show that BPS preschool boosts postsecondary education for a population with high shares of minority and low-income children: 72% are Black or Hispanic, and more than two-thirds are eligible for a free or reduced-price lunch (a proxy for low family income).[19] We next turn to an analysis of outcomes prior to college to investigate the channels driving these results.

V. Effects on College Preparation

V.A. *Grade Progression, Special Education, and High School Graduation*

Studies of preschool often consider outcomes related to grade retention and special education status (Gramlich 1986; Currie and Thomas 1995; Currie 2001; Magnuson et al. 2004; Deming 2009; Miller and Bassok 2017). Preschool may ease the transition to elementary school and reduce the need for remediation and special education services (Bailey et al. 2017). Currie (2001) emphasizes that prevention of special education and avoidance of grade retention are potential cost savings created by preschool programs. Furthermore, special education classification and grade progression outcomes may contain information on skills and behaviors that are not captured by test scores.

We find no detectable effects of BPS preschool on grade repetition and special education outcomes. The first row of Table IV displays a 2SLS estimate of the effect of BPS preschool on starting first grade on time. The sample for this outcome is limited to those who are observed in first grade in a Massachusetts public school at some point. Eighty-seven percent of nonoffered students start first grade on time, and the estimated effect of BPS preschool is

19. The fraction of students eligible for a subsidized lunch is calculated based on free or reduced-price lunch status in the first year a student appears in the SIMS database using students who appear in the SIMS data in at least one year.

TABLE IV

EFFECTS OF PRESCHOOL ATTENDANCE ON GRADE PROGRESSION, SPECIAL EDUCATION, AND HIGH SCHOOL GRADUATION

	Nonoffered mean (1)	2SLS estimate (2)
Panel A: Grade progression and special education		
Started 1st grade on time	0.874	0.016
		(0.023)
	1,529	2,375
Enrolled in BPS in 6th grade	0.810	0.032
		(0.024)
	2,459	3,883
Enrolled in BPS in 9th grade	0.806	0.031
		(0.025)
	2,459	3,883
Repeated a grade	0.325	−0.036
		(0.029)
	2,459	3,883
Special education in 1st grade	0.090	0.012
		(0.021)
	1,529	2,375
Special education in 3rd grade	0.144	0.003
		(0.022)
	2,459	3,883
Panel B: High school graduation		
Graduated high school on time	0.624	0.054*
		(0.030)
	2,459	3,883
Ever graduated high school	0.636	0.060**
		(0.030)
	2,459	3,883

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on grade progression, special education classification, and high school graduation. The on-time first-grade outcome equals one if a student appears in first grade by the expected year given their BPS preschool application year. The sample for this outcome is restricted to students observed in first grade in a Massachusetts public school. BPS enrollment outcomes equal one if a student is ever observed enrolled in a BPS school for the relevant grade. The grade repetition outcome equals one for students who are ever observed in the same grade in multiple years. Samples for BPS enrollment and grade repetition include students who are ever observed in a Massachusetts public school. Samples for first- and third-grade special education include students observed in the relevant grade. On-time high school graduation equals one if a student is recorded as graduating from a Massachusetts public high school by the end of his/her projected 12th-grade year. Samples for the graduation outcomes include students who are ever observed in a Massachusetts public school. Column (1) displays the nonoffered mean for each outcome. Column (2) reports 2SLS coefficients from models instrumenting preschool attendance with the preschool offer. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

small and statistically insignificant. Similarly, we find no effect on the probability of appearing in a BPS school in sixth or ninth grade, and a small negative but statistically insignificant effect on repeating a grade (defined as appearing in the same grade in more than one year). The samples for these outcomes are restricted to students who appear in a Massachusetts public school in at least one year. The bottom rows of Table IV, Panel A show small and insignificant estimates of effects on special-education classification in first and third grades for students observed in these grades. Online Appendix Table A11 shows a similar pattern of small and insignificant grade repetition and special-education effects across all grades between kindergarten and grade 12.

In contrast, Table IV, Panel B reveals that preschool attendance boosts high school graduation. Enrollment in a BPS preschool increases the probability that students graduate from a Massachusetts public high school on time by 5.4 percentage points ($p < .1$), and this effect grows to 6.0 percentage points when we include students who graduate at any time ($p < .05$). The estimated effect on ever graduating high school is a 9.4% increase relative to the nonoffered graduation rate of 64%. Combined with the insignificant effects on grade repetition in Panel A, the high school graduation results suggest that BPS preschool increases the likelihood that students successfully complete high school rather than causing them to enroll in high school earlier. It's worth emphasizing, however, that these estimates are based on students who appear in the Massachusetts public school database at some point, so the results should be interpreted with some caution given the differential attrition documented in Table II.

V.B. SAT Test Taking and Scores

The SAT is an important assessment for college-bound high school students because it is widely used for college admissions. Students usually take the SAT in 11th or 12th grade after taking standardized tests in Massachusetts required for high school graduation. The SAT outcome is also of particular interest for the low-income population studied here, since the SAT is seen as a significant hurdle for students who may not have access to test preparation (see Bowen and Bok 2000).

Enrollment in a BPS preschool raises the likelihood that students take the SAT. Table V, column (1) shows that among nonoffered BPS preschool applicants who attend a Massachusetts public high school, roughly two-thirds take the SAT. Preschool attendance causes a statistically significant 8.5 percentage point increase in the rate of SAT test taking. The size of this effect is similar to the estimated effect of preschool attendance on on-time

TABLE V

EFFECTS OF PRESCHOOL ATTENDANCE ON SAT TEST TAKING AND SCORES

	Taking		Reasoning (1600)		Composite (2400)	
	Nonoffered mean (1)	2SLS estimate (2)	Nonoffered mean (3)	2SLS estimate (4)	Nonoffered mean (5)	2SLS estimate (6)
Took SAT	0.685	0.085**				
		(0.034)				
Score above MA bottom quartile			0.384	0.062*	0.376	0.057
				(0.036)		(0.035)
Score above MA median			0.225	0.008	0.216	– 0.001
				(0.030)		(0.029)
Score in MA top quartile			0.096	0.036*	0.092	0.031
				(0.021)		(0.021)
N						2,559
Average score			940.6	3.9	1,392.0	7.2
(for takers)				(16.6)		(24.2)
N						1,863

TABLE V
CONTINUED

	Math (800)		Verbal (800)		Writing (800)	
	Nonoffered mean (1)	2SLS estimate (2)	Nonoffered mean (3)	2SLS estimate (4)	Nonoffered mean (5)	2SLS estimate (6)
Score above MA bottom quartile	0.419	0.070*	0.366	0.063*	0.379	0.050
		(0.037)		(0.036)		(0.036)
Score above MA median	0.241	0.002	0.202	0.007	0.205	0.057*
		(0.030)		(0.029)		(0.030)
Score in MA top quartile	0.097	0.057***	0.103	0.002	0.087	0.004
		(0.021)		(0.022)		(0.020)
N						2,559
Average score	482.5	−0.6	458.1	4.5	451.4	3.3
(for takers)		(8.8)		(9.1)		(8.7)
N						1,863

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on SAT test taking and scores. The sample is restricted to students with a 10th-grade MCAS score. Outcomes for scoring above the bottom quartile, above the median, and in the top quartile are coded to zero for students who did not take the SAT. Column (1) displays mean outcomes for nonoffered students, and column (2) displays 2SLS estimates instrumenting BPS preschool attendance with the preschool offer. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

college enrollment, suggesting that taking the SAT may play a role in accounting for our results on increased college attendance.

Because preschool attendance affects SAT test taking, we examine how preschool influences SAT performance defined by unconditional score thresholds. A student scores above a given quartile in the state distribution if she takes the SAT and scores above the threshold defined by the state quartile. Students who do not take the test and those who take the test but fail to reach the relevant threshold are coded as zeros for these outcomes. Table V, column (3) shows that by this definition, less than one-quarter of preschool applicants score above the state median on the SAT Reasoning test (defined as the sum of Math and Verbal scores). Likewise, column (5) shows that less than one-quarter score above the state median on the SAT Composite test (defined as the sum of Math, Verbal, and Writing).

Preschool attendance affects the bottom quartile of performance on the SAT. Table V, column (4) shows that preschool attendance marginally increases the likelihood that a student scores above the bottom quartile on SAT Reasoning. Because the outcome is coded as zero for nontakers, this effect combines the extensive-margin impact on SAT-taking and any intensive-margin effects on scores. The effect on scores in the bottom quartile is driven by a 5–7 percentage point improvement in the likelihood of clearing the bottom quartile in each component subject test, with the estimates for Math and Verbal significant at the 10% level.

Estimates for the top quartile suggest that preschool attendance increases the likelihood of achieving a high SAT math score. We find a statistically significant increase of 5.7 percentage points in the share of students scoring in the top quartile in math, a large effect relative to the control mean of 9.7%. The estimated effect on top-quartile Reasoning and Composite scores are also positive, though only the Reasoning estimate is marginally significant. The bottom row of Table V also reports 2SLS estimates of effects on average SAT scores in the sample of test takers. These conditional results are difficult to interpret since preschool offers have a large impact on the likelihood of taking the test. Estimated effects on average SAT scores are imprecise and statistically indistinguishable from zero, but these estimates may be contaminated by composition effects because of the impact of preschool attendance on SAT test taking. Taken at face value, the estimated effects among takers indicate little effect on SAT scores, suggesting that the effects we see on score thresholds are likely to be driven by the increased likelihood of taking the test.

VI. Effects on Test Scores and Disciplinary Outcomes

VI.A. MCAS Scores

Previous studies of Boston preschools show that for recent applicant cohorts, the program increased test scores measured during the preschool year, but test score effects were not detectable by third grade (Weiland and Yoshikawa 2013; Weiland et al. 2019). We do not observe test scores in the preschool year for our applicant sample, but we can study effects on medium-term test scores on the MCAS. Massachusetts started administering MCAS exams in 1998 with tests in grade 4 and 8. The state subsequently expanded tests to other grades, and tests are now administered in grades 3–8 and 10. MCAS performance is consequential for schools, since it factors into the state's accountability framework. A student must also pass MCAS Math and English Language Arts (ELA) tests to earn a high school diploma.

Table VI reports estimated effects of preschool attendance on MCAS test scores. We standardize these scores to have mean zero and standard deviation one in the sample of all Massachusetts test takers in each grade and year. Among nonoffered BPS preschool applicants, mean scores on Math and ELA tests in elementary school are around -0.3σ to -0.4σ, implying achievement substantially below the state average. As shown in columns (2) and (4), we find that preschool attendance has no statistically detectable impact on these achievement levels. This result is consistent with Weiland et al. (2019) who report that attendance at a first-choice BPS preschool did not affect third-grade MCAS performance for cohorts applying between 2007 and 2011. More broadly, our findings echo those in other recent randomized studies of preschool programs, which often find limited achievement effects in elementary school (Puma et al. 2010; Lipsey, Farran, and Durkin 2018).

Our data offer the opportunity to study achievement effects in middle and high school as well. As in elementary school, we find no evidence of effects on MCAS scores in later grades. Table VI shows a mix of positive and negative estimates for grades 6, 7, 8, and 10. None of the estimates are significantly different from zero. The bottom row shows 2SLS estimates from a model that stacks all observed MCAS scores in grades 3–8 and 10, with

TABLE VI
EFFECTS OF PRESCHOOL ATTENDANCE ON MCAS TEST SCORES

	Math scores		ELA scores	
	Nonoffered mean (1)	2SLS (2)	Nonoffered mean (3)	2SLS (4)
Grade 3	−0.400	0.024	−0.424	−0.048
		(0.094)		(0.068)
	677	1,092	2,025	3,241
Grade 4	−0.302	−0.063	−0.339	−0.025
		(0.066)		(0.067)
	2,022	3,226	2,020	3,219
Grade 5	−0.276	0.022	−0.366	0.071
		(0.076)		(0.080)
	1,319	2,059	1,316	2,056
Grade 6	−0.221	−0.023	−0.311	0.027
		(0.067)		(0.072)
	1,948	3,113	1,690	2,625
Grade 7	−0.180	−0.003	−0.203	0.049
		(0.064)		(0.064)
	1,950	3,114	1,948	3,109
Grade 8	−0.157	0.024	−0.194	−0.009
		(0.063)		(0.065)
	1,939	3,093	1,936	3,087
Grade 10	−0.096	−0.031	−0.158	0.066
		(0.064)		(0.062)
	1,785	2,847	1,801	2,852
All grades (stacked)	−0.215	0.005	−0.283	0.029
		(0.057)		(0.056)
Number of students	2,249	3,569	2,279	3,615
Number of scores	11,640	18,544	12,736	20,189

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on Massachusetts Comprehensive Assessment System (MCAS) achievement test scores. MCAS scores are standardized to have mean zero and standard deviation one among all Massachusetts test-takers. The bottom row stacks all observed test scores in grades 3–8 and 10, and clusters standard errors by student. Columns (1) and (2) show results for math scores, and columns (3) and (4) show results for English Language Arts (ELA) scores. Columns (1) and (3) display mean outcomes for nonoffered students. Columns (2) and (4) show 2SLS coefficients from models instrumenting preschool attendance with the preschool offer. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

standard errors clustered by student. Estimates from this model are also statistically insignificant, and the precision of the estimates allows us to rule out positive effects larger than 0.12σ in Math and 0.14σ in ELA. Although we observe at least one MCAS score for roughly 85% of the sample and an average of nine test scores per nonoffered applicant, an important caution is that the

MCAS follow-up differential shown in Table II may influence our estimates of test score effects.

VI.B. Disciplinary Outcomes

While the results of the previous section suggest that BPS preschool has limited effects on test scores, previous studies suggest that preschool programs can generate persistent effects through noncognitive channels. For example, Heckman et al. (2010b) emphasizes the role of criminal justice outcomes in the high social rate of return to the Perry Preschool program. Heckman, Pinto, and Savelyev (2013) demonstrate that the Perry intervention substantially improved externalizing behaviors (aggressive, antisocial, and rule-breaking behaviors), and that these effects account for the bulk of its long-term effects. A related literature shows effects of teachers and schools on nontest outcomes such as suspensions, truancy, absenteeism, course grades, and crime (Deming 2011; Jackson 2018; Petek and Pope 2021). Bacher-Hicks, Billings, and Deming (2019) and Rose, Schellenberg, and Shem-Tov (2022) argue that short-term effects of teachers and schools on noncognitive outcomes predict longer-term effects on criminal behavior. Using a regression discontinuity design, Weiland and Yoshikawa (2013) estimate short-term positive effects of BPS preschools on executive function and a measure of emotion recognition.

Motivated by these findings, Table VII displays 2SLS estimates of the effects of preschool attendance on several disciplinary outcomes measured in middle and high school. We look at effects on suspensions, truancy, absences, and juvenile incarceration. We measure juvenile incarceration based on whether a student is ever observed attending a Massachusetts Department of Youth Services (DYS) school.[20] To aggregate the potentially noisy effects on individual outcomes, the bottom row of Table VII reports estimated effects on a summary index of discipline. Following Jackson (2018), this index is formed as the first principal component of these outcomes, scaled to have mean zero and standard deviation one among nonoffered students (and defined so that a positive estimate means a decline in disciplinary problems). Column (2)

20. DYS operates the state's juvenile justice service, and DYS facilities provide rehabilitation for students who have committed crimes. This measure seems to be a reliable measure of incarceration, as we never observe a student simultaneously enrolled in a traditional public school and a DYS facility.

TABLE VII

Effects of Preschool Attendance on Disciplinary Outcomes

	Middle school		High school	
	Nonoffered mean (1)	2SLS (2)	Nonoffered mean (3)	2SLS (4)
Ever suspended	0.171	−0.012	0.166	−0.021
		(0.024)		(0.023)
	2,208	3,512	2,099	3,335
Number of suspensions	0.593	−0.131	0.663	−0.241*
		(0.150)		(0.141)
	2,208	3,512	2,099	3,335
Ever truant	0.167	0.004	0.654	0.027
		(0.021)		(0.029)
	2,208	3,512	2,099	3,335
Times truant	0.695	0.001	28.05	−4.408
		(0.218)		(3.557)
	2,208	3,512	2,099	3,335
Days absent	8.380	0.100	66.20	−6.467
		(0.580)		(5.047)
	2,208	3,512	2,099	3,335
Juvenile incarceration	0.001	−0.001	0.007	−0.010**
		(0.001)		(0.005)
	2,208	3,512	2,099	3,335
Disciplinary index	0.000	−0.008	0.000	0.172***
		(0.064)		(0.062)
	2,208	3,512	2,099	3,335

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on disciplinary outcomes measured in middle school (grades 6–8) and high school (grades 9–12). The sample is restricted to students observed in a Massachusetts public school in ninth grade. Juvenile incarceration equals one if a student is ever recorded as incarcerated or attending a Department of Youth Services institution. The noncognitive index is the first principle component of all outcomes in the table, standardized to have mean zero and standard deviation one among nonoffered students. Column (1) displays the nonoffered mean for each outcome, and column (2) reports coefficients from 2SLS models instrumenting preschool attendance with the preschool offer. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

shows no effect on disciplinary outcomes in middle school, with insignificant estimates for all individual outcomes as well as the summary index in the bottom row.

Estimates for high school, shown in column (4), show that preschool attendance reduces the likelihood of disciplinary problems in high school. We see a marginally significant decline in the number of times students are suspended ($p < .1$). Only 1% of nonoffered preschool applicants are ever incarcerated according to our measure, but preschool enrollment is estimated to reduce juvenile incarceration by 1 percentage point ($p < .05$). Although estimates for the individual outcomes are imprecise, we find that preschool attendance boosts the high school discipline index by 0.17σ, a large and statistically significant effect ($p < .01$).

VII. Further Results and Discussion

VII.A. Effects on Subgroups

The literature on early childhood programs often finds important differences in treatment effects across student subgroups. In a reanalysis of the Abecedarian, Perry, and Early Training Projects, Anderson (2008) finds significant short- and long-term benefits for girls but no significant long-term effects for boys after adjusting for multiple testing. Heckman et al. (2010a) account for compromised randomization in the Perry experiment and find that it generated significant long-term benefits for both sexes. Magnuson et al. (2016) provide an overview of research on gender differences in preschool effects. Research on Head Start has emphasized differences in effects between Black, white, and Hispanic students (Currie and Thomas 1995; Garces, Thomas, and Currie 2002; Deming 2009; Bitler, Hoynes, and Domina 2014; Bloom and Weiland 2015). Gormley and Gayer (2005) show that Tulsa's public preschool program produces larger gains for middle-income students than for low-income students. Cascio (forthcoming) finds larger test score effects for universal state-funded preschool programs than for means-tested programs, including larger effects on low-income students. Weiland and Yoshikawa (2013) report larger short-term effects of Boston preschools for lower-income and Latino students in recent cohorts.

Table VIII reports 2SLS estimates of the effects of BPS preschool attendance on key outcomes for student subgroups. We probe for effect heterogeneity by sex, race, and income. Since large differences in point estimates are likely to arise by chance with small subgroup samples, we report p-values from tests of the hypothesis that effects are equal for each sample split. We also show p-values from tests of the null hypothesis that there are no differences for any outcome for a given sample split, as well as tests that estimates are equal across all subgroups for a given outcome.

The effects of BPS preschool attendance are generally larger for boys than for girls. As shown in Table VIII, columns (1) and (2), preschool enrollment increases on-time college enrollment for both sexes, but effects on four-year college enrollment are driven

TABLE VIII

EFFECTS OF PRESCHOOL ATTENDANCE FOR SUBGROUPS

	By sex		By race			By free/reduced-price lunch		Joint p-value across subgroups
	Boys (1)	Girls (2)	Black (3)	Hispanic (4)	White (5)	FRPL (6)	Not FRPL (7)	(8)
Any college enrollment (on-time)	0.117**	0.076*	0.032	0.134***	0.214**	0.069*	0.134**	.254
	(0.046)	(0.042)	(0.047)	(0.048)	(0.100)	(0.039)	(0.055)	
	2,099	2,076	1,692	1,405	643	2,510	1,373	
p-value for equal effects	.510			.136		.333		
Four-year college enrollment (on-time)	0.137***	0.043	0.056	0.092**	0.254**	0.066*	0.149***	.124
	(0.045)	(0.041)	(0.045)	(0.046)	(0.102)	(0.037)	(0.055)	
	2,099	2,076	1,692	1,405	643	2,510	1,373	
p-value for equal effects	.122			.193		.210		
Any college enrollment (ever)	0.074	0.053	0.040	0.069	0.159*	0.026	0.102**	.536
	(0.045)	(0.039)	(0.046)	(0.048)	(0.089)	(0.037)	(0.049)	
	2,099	2,076	1,692	1,405	643	2,510	1,373	
p-value for equal effects	.734			.471		.218		
Four-year college enrollment (ever)	0.126***	−0.006	0.059	0.041	0.211**	0.024	0.136***	.038
	(0.046)	(0.042)	(0.047)	(0.049)	(0.096)	(0.039)	(0.053)	
	2,099	2,076	1,692	1,405	643	2,510	1,373	
p-value for equal effects	.0339			.257		.083		
Ever graduated high school	0.163***	−0.035	0.066	0.009	0.161*	0.064*	0.102**	.008
	(0.046)	(0.040)	(0.047)	(0.049)	(0.090)	(0.038)	(0.052)	
	1,899	1,861	1,526	1,268	552	2,433	1,319	
p-value for equal effects	.00147			.175		.557		

TABLE VIII
CONTINUED

	By sex		By race			By free/reduced-price lunch		Joint p-value across subgroups
	Boys (1)	Girls (2)	Black (3)	Hispanic (4)	White (5)	FRPL (6)	Not FRPL (7)	(8)
MCAS math scores (stacked)	0.070*	−0.064*	0.056	−0.131***	0.195**	−0.075**	0.175***	.100
	(0.041)	(0.036)	(0.038)	(0.040)	(0.088)	(0.031)	(0.050)	
	1,797	1,760	1,446	1,194	539	2,351	1,206	
p-value for equal effects	.226			.112		.038		
Took SAT	0.158***	0.028	0.132**	0.068	0.028	0.092**	0.106*	.333
	(0.055)	(0.045)	(0.052)	(0.062)	(0.094)	(0.044)	(0.056)	
	1,247	1,260	1,015	791	371	1,650	852	
p-value for equal effects	.069			.540		.841		
Disciplinary index	0.344***	0.035	0.238**	0.118	0.090	0.235***	0.107	.106
	(0.106)	(0.074)	(0.107)	(0.093)	(0.169)	(0.081)	(0.100)	
	1,616	1,597	1,296	1,067	460	2,129	1,086	
p-value for equal effects	.023			.632		.319		
Joint p-value across outcomes	.022			.098		.200		

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on key outcomes for subgroups. Columns (1) and (2) compare estimates for boys and girls, columns (3)–(5) display estimates by race, and columns (6) and (7) show estimates by free/reduced-price lunch (FRPL) status. Students are included in the FRPL analysis if they appear in the SIMS database, and are classified as FRPL if they receive a subsidized lunch in their first year in the SIMS. p-values for equal effects come from tests of the null hypothesis that effects are equal across subgroups. Estimates are 2SLS coefficients instrumenting preschool attendance with the preschool offer. All models control for a saturated set of indicators for the assignment propensity score. Column (8) shows joint p-values from tests of the null hypothesis of no heterogeneity in effects across any of the sample splits for the relevant outcome. The last row shows joint p-values from tests of the null hypothesis that subgroups have equal effects across all outcomes in the table.
* significant at 10%; ** significant at 5%; *** significant at 1%.

by boys. Similarly, for boys we find significant increases in SAT-taking (16 percentage points), high school graduation (16 percentage points), and the discipline index (0.34σ), with small and insignificant corresponding estimates for girls. A joint test rejects the null hypotheses of equal effects for boys and girls across all outcomes ($p = .02$).

Differences in estimates by race and income are mostly statistically insignificant. Table VIII, columns (3), (4), and (5) present estimates for Black, Hispanic, and white students. The pattern of point estimates suggests somewhat larger effects for whites, but the white subgroup is small, and none of the racial differences in estimates are statistically significant at conventional levels ($p > .11$), though a joint test rejects the null hypothesis of no heterogeneity across any outcome at marginal significance levels ($p = .10$). To assess heterogeneity by income, columns (6) and (7) show estimates for students eligible and ineligible for a free or reduced-price lunch, a proxy for low family income. Students are included in this analysis if they appear in the SIMS database and are classified as eligible for a subsidized lunch if they are recorded as receiving a free or reduced-price lunch in their first year in the SIMS. Estimates for low- and higher-income students are generally not statistically distinguishable (joint $p = .20$).

Differences by subgroup need not solely be driven by differences in preschool effectiveness. They could also be driven by differences in fallback options or differences in the types of public preschools attended by subgroup (for a related discussion, see Angrist, Pathak, and Walters 2013). For example, low-income students may be particularly likely to attend Head Start programs if they do not win a seat at Boston preschool. Without direct measurement of the counterfactual state, we are unable to draw firm conclusions about the sources of potential subgroup differences.

VII.B. School and Peer Characteristics

Previous research suggests that the quality of primary schools may play a role in fadeout of preschool test score impacts (Currie and Thomas 1995; Johnson and Jackson 2019; Weiland et al. 2019). We assess this possibility by estimating the test score value-added of schools attended by BPS preschool applicants. Value-added is calculated by regressing MCAS math scores on a set of school indicators with controls for lagged test scores and demographics.[21] We run these regressions separately by grade for the full population of Massachusetts test takers during our study period. As shown in Table IX, column (1), attending a BPS preschool has no detectable effect on the value-added of the schools that a student subsequently attends. This result indicates that changes in overall school quality are unlikely to mediate the causal effects of BPS preschools.

Preschool attendance may affect students' K–12 school trajectories on dimensions other than test score value-added. Columns (2) and (3) report effects on the probability of attendance at charter and exam schools. We find some suggestive evidence that BPS preschool reduces the likelihood of attending a charter school in early grades, perhaps because attending a district preschool increases attachment to traditional district schools. Estimated effects on charter and exam attendance in grades 7–12 are small and statistically insignificant.

Columns (4)–(6) investigate whether preschool shifts school attendance in a manner that alters the composition of students' peer groups, measured using the average characteristics of other students attending the same school, grade, and year. Column (4) demonstrates that preschool attendance has no consistent effect on peer math achievement. In contrast, column (5) demonstrates that BPS preschool attendance increases the share of a student's grade K–6 peers that will eventually go on to attend college. Attending a BPS preschool also boosts the likelihood of attending school with classmates who themselves enrolled in BPS preschool, as can be seen in column (6). This increase in BPS preschool peer share likely stems from students remaining in their BPS preschools into elementary school (some schools serve both pre-K and higher grades). The increased peer college-going rate in column (5) may reflect the positive effect of BPS preschool on college-going combined with an increased likelihood of attending school with BPS preschool peers. Taken together, the results in Table IX provide suggestive evidence that changes in school attendance and peer composition may play a role in the effects of preschool

21. Following recent work on teacher and school value-added (Chetty, Friedman, and Rockoff 2014a; Angrist et al. 2017), the value-added models control for race, sex, subsidized lunch status, limited English proficiency, lagged absences and suspensions, and cubic functions of Math and ELA scores from the most recent available grade. The value-added for a particular school is estimated by the coefficient on the corresponding school indicator.

TABLE IX

EFFECTS OF PRESCHOOL ATTENDANCE ON SCHOOL QUALITY, SCHOOL CHOICE, AND PEER CHARACTERISTICS

	School quality and choice			Peer characteristics		
	Test score value-added (1)	Attends charter school (2)	Attends exam school (3)	Avg. math score (4)	College attendance (5)	BPS preschool attendance (6)
Grade K		−0.019			0.056***	0.277***
		(0.014)			(0.007)	(0.020)
		1,618			1,544	1,544
Grade 1		−0.021*			0.057***	0.151***
		(0.013)			(0.007)	(0.014)
		2,189			2,117	2,117
Grade 2		−0.024**			0.027***	0.071***
		(0.012)			(0.008)	(0.010)
		2,852			2,777	2,777
Grade 3		−0.021*			0.030***	0.054***
		(0.012)			(0.008)	(0.008)
		3,385			3,318	3,318
Grade 4	0.019	−0.025**		0.033	0.022***	0.040***
	(0.016)	(0.012)		(0.034)	(0.008)	(0.007)
	3,323	3,352		3,221	3,298	3,298
Grade 5	0.028**	−0.004		0.090**	0.032***	0.028***
	(0.014)	(0.017)		(0.037)	(0.009)	(0.007)
	3,299	3,319		2,057	3,243	3,243
Grade 6	0.007	−0.042*		0.033	0.021**	0.027***
	(0.015)	(0.022)		(0.036)	(0.009)	(0.006)
	3,253	3,258		3,111	3,183	3,183
Grade 7	−0.011	−0.022	−0.018	0.019	0.014	0.020***
	(0.017)	(0.021)	(0.023)	(0.042)	(0.011)	(0.006)
	3,254	3,261	3,261	3,109	3,192	3,192
Grade 8	−0.000	−0.016	−0.001	0.027	0.006	0.017***
	(0.015)	(0.021)	(0.023)	(0.041)	(0.011)	(0.005)
	3,187	3,193	3,193	3,086	3,171	3,171
Grade 9		−0.017	0.006		−0.009	0.002
		(0.022)	(0.025)		(0.013)	(0.002)
		3,237	3,237		3,151	3,151
Grade 10	0.008	0.006	0.019	0.006	−0.001	0.002
	(0.019)	(0.021)	(0.025)	(0.045)	(0.013)	(0.002)
	3,069	3,070	3,070	2,829	3,022	3,022

TABLE IX
CONTINUED

	School quality and choice			Peer characteristics		
	Test score value-added (1)	Attends charter school (2)	Attends exam school (3)	Avg. math score (4)	College attendance (5)	BPS preschool attendance (6)
Grade 11		0.003	0.013		0.007	0.002
		(0.022)	(0.026)		(0.013)	(0.002)
		2,915	2,915		2,876	2,876
Grade 12		−0.002	0.017		0.016	0.004
		(0.022)	(0.026)		(0.012)	(0.003)
		2,843	2,843		2,721	2,721

Notes. This table reports two-stage least squares (2SLS) estimates of the effects of Boston preschool attendance on school quality and choice, and peer characteristics. All columns report 2SLS coefficients models instrumenting preschool attendance with the preschool offer. Value-added measures are obtained by estimating OLS regressions of MCAS math scores on school enrollment indicators with controls for cubic functions of lagged Math and ELA scores, sex, race, free/reduced lunch status, special education status, English language learner status, lagged absences and suspensions, and year indicators. Value-added regressions are estimated separately by grade using all Massachussetts students in the cohorts listed in Online Appendix Table A1. Value-added is normalized to mean zero in traditional BPS schools by subtracting the BPS student-weighted average in each grade. Preschool applicants are then assigned the estimated value-added for the school they attend in each grade. School peer composition characteristics in columns (4)–(6) are constructed by taking the average of that characteristic for all peers who attended the same school, grade, and year. All models control for a saturated set of indicators for the assignment propensity score along with sex, bilingual status, and race. * significant at 10%; ** significant at 5%; *** significant at 1%.

attendance, although we are not able to separately parse these mechanisms.

VII.C. Comparison to Estimates in the Literature

Previous estimates of the effects of preschool programs on educational attainment come from small-scale experiments and nonexperimental studies of the Head Start program for earlier cohorts. To understand whether the effects of large-scale preschool in Boston differ from effects for these earlier programs, we compare our results to estimates from prominent studies in the literature. Table X lists study characteristics and estimated educational attainment effects for our evaluation of the Boston preschool program along with several previous studies evaluating other programs, including the Perry Preschool Project, Abecedarian Project, and Head Start.[22]

22. Some subsequent analyses building on the studies in Table X arrive at different estimates for the same programs due to changes in sample or methodology

TABLE X

COMPARISON OF STUDY CHARACTERISTICS AND EDUCATIONAL ATTAINMENT ESTIMATES FROM PRESCHOOL EVALUATIONS

					Impact estimates (std. err.)	
Study	Program	Randomized design (1)	Long-term outcomes (2)	Large-scale program (3)	High school grad. (4)	College attendance (5)
This article	Boston preschool	Yes	Yes	Yes	0.060 (0.030)	0.054 (0.029)
Belfield et al. (2006)[a]	Perry Preschool Project	Yes	Yes	No	0.165 (0.084)	–
Campbell et al. (2012)[b]	Abecedarian Project	Yes	Yes	No	0.068 (0.072)	0.170 (0.068)
Garces et al. (2002)[c]	Head Start	No	Yes	Yes	0.037 (0.053)	0.092 (0.056)
Deming (2009)[d]	Head Start	No	Yes	Yes	0.086 (0.031)	0.057 (0.036)
Bailey et al. (2021)[e]	Head Start	No	Yes	Yes	0.024 (0.012)	0.054 (0.028)
Puma et al. (2010)	Head Start	Yes	No	Yes	–	–
Lipsey et al. (2018)	TN-VPK	Yes	No	Yes	–	–
Weiland et al. (2019)	Boston preschool	Yes	No	Yes	–	–
				Average effect:	0.038 (0.007)	0.064 (0.012)
			p-value for test of no heterogeneity:		.239	.564

Notes. This table compares study characteristics and educational attainment estimates from prominent evaluations of preschool programs. The top row shows estimates for the Boston preschool program studied in this article, and the remaining rows display literature studies of other programs. Column (1) labels studies that use research designs with random assignment, column (2) labels studies that look at long-term outcomes, and column (3) labels studies that evaluate large-scale programs. For evaluations with long-term outcomes, column (4) shows each program's estimated effect on ever graduating high school and the corresponding standard error, and column (5) reports estimates and standard errors for the effect of attending college. Average effects and p-values come from a classical minimum distance (CMD) procedure that treats each estimate in a column as an independent estimate of the same effect. The average effect is the CMD estimate and the p-value comes from a comparison of the minimized CMD criterion function to a chi-squared distribution with degrees of freedom equal to the number of estimates minus one.

[a] High school graduation estimate and standard error are derived from the age 40 counts and percentages in Table 1. College impact is not reported because few students attended college.

[b] Estimates and standard errors are derived from counts and percentages in Table 3. The estimate in column (5) is the impact on graduation because the effect on any attendance is not reported.

[c] Mother fixed effects estimates from Table 2.

[d] Family fixed effects estimates from Table 5.

[e] ATET estimates from Table 1. Standard errors are calculated as the width of the 95% confidence interval divided by 3.92.

Three key patterns are evident in this comparison. First, as shown in columns (1)–(3), ours is the only study that combines a randomized design, long-term outcomes, and a large-scale program. The other studies listed in Table X lack one of these characteristics. It is important to note, however, that several of the other studies are able to look at other long-term outcomes, such as earnings and criminal activity, whereas we can only look at educational attainment. Second, the standard errors in columns (4) and (5) show that the precision of our design compares favorably to most previous studies. The precision of our postsecondary effect estimates is comparable to estimates from Bailey, Timpe, and Sun (2021)'s study of the initial rollout of Head Start using the Social Security Administration Numident file (though our estimates for high school graduation are less precise). Third, our effect estimates for educational attainment are consistent with estimates from previous studies. Though the studies in Table X estimate a variety of parameters for multiple programs using a mix of randomized and nonrandomized research designs, the estimated effects on high school graduation and college enrollment are surprisingly similar.

The bottom rows of Table X formally investigate the similarity of effect sizes across studies. Specifically, we use classical minimum distance (CMD) to fit a model that assumes the effect in each study is the same, treating each study as an independent unbiased estimate of this single effect with variance equal to its squared standard error. Under the null hypothesis of no heterogeneity in effects across studies, the minimized CMD criterion function follows a χ^2 distribution with degrees of freedom equal to the number of studies minus one. This CMD procedure generates precise average effect estimates of 3.8 percentage points for high school graduation (std. err. = 0.7) and 6.4 percentage points for college attendance (std. err. = 1.2). The χ^2 goodness of fit test fails to reject for either outcome ($p > .23$), indicating that the differences in estimates across studies can be rationalized by sampling error. This exercise reveals a consistent picture of positive preschool effects on educational attainment across a diverse

(see Heckman et al. 2010a; Pages et al. 2020; Miller et al. forthcoming). We include studies in Table X to provide one set of benchmark estimates against which our estimates for Boston preschool can be compared.

set of studies. Our results show that this pattern continues to hold in a randomized evaluation of a large-scale public preschool program.

VIII. Conclusion

High-quality preschool programs have the potential to produce lasting effects on skills and improve long-term outcomes for disadvantaged students (Elango et al. 2016). While public preschool has expanded rapidly in recent decades, little evidence exists on the long-term impacts of modern large-scale preschool programs. Such evidence is important both for understanding the efficacy of programs operating at scale and for interpreting links between short-term and long-term outcomes. This article uses random variation from Boston's centralized school assignment mechanism to provide the first evidence on the long-term effects of a modern, large-scale public preschool program from a lottery-based research design.

The results of our analysis show that public preschool enrollment boosts postsecondary and college preparatory outcomes. Students randomly assigned to attend a Boston preschool experience fewer disciplinary incidents in high school, take the SAT and graduate high school at higher rates, and are more likely to enroll in college. These findings illustrate the potential for a universal preschool program to improve educational attainment for a disadvantaged student population.

Boston's public preschool program expanded rapidly after the time period of our study, and the district made efforts to improve preschool quality (Sachs and Weiland 2010; Weiland et al. 2013). It is therefore possible that the effects reported here differ from effects for more recent cohorts in Boston or for lower-quality programs elsewhere. At the same time, Boston's program shares important features with other publicly funded state and local preschool programs, so our estimates seem relevant for evaluating contemporary proposals for public preschool expansion (Biden 2021). Although we are able to document effects on educational attainment, other work has shown effects of early childhood programs on even longer-term outcomes such as employment, earnings, and criminal activity (Garcia et al. 2020). In future work, we hope to study effects on these and other economic outcomes over the life cycle.

University of Chicago, United States
Massachusetts Institute of Technology and National Bureau of Economic Research, United States
University of California, Berkeley, and National Bureau of Economic Research, United States

Supplementary Material

Supplementary material is available at the *Quarterly Journal of Economics* online.

Data Availability

Code replicating the tables and figures in this article can be found in Gray-Lobe, Pathak, and Walters (2022) in the Harvard Dataverse, https://doi.org/10.7910/DVN/BOU7XR.

References

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, and Thomas J. Kane, "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *Quarterly Journal of Economics*, 126 (2011), 699–748.
Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 85 (2017), 1373–1432.
Abdulkadiroğlu, Atila, and Tayfun Sönmez, "School Choice: A Mechanism Design Approach," *American Economic Review*, 93 (2003), 729–747.
Almond, Douglas, Janet Currie, and Valentina Duque, "Childhood Circumstances and Adult Outcomes: Act II," *Journal of Economic Literature*, 56 (2018), 1360–1446.
Anders, John, Andrew Barr, and Alexander A. Smith, "The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s," *American Economic Journal: Economic Policy* (forthcoming).
Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103 (2008), 1481–1495.
Angrist, Joshua, Parag A. Pathak, and Christopher R. Walters, "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 5 (2013), 1–27.
Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132 (2017), 871–919.
Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91 (1996), 444–455.
Armor, David J., "The Evidence on Universal Preschool: Are Benefits Worth the Cost?," CATO Institute: Policy Analysis, 2014.
Athey, Susan, Raj Chetty, Guido Imbens, and Hyunseung Kang, "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," NBER Working Paper no. 26463, 2019.

Bacher-Hicks, Andres, Stephen B. Billings, and David J. Deming, "The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime," NBER Working Paper 26257, 2019.
Bailey, D., G. Duncan, C. Odgers, and W. Yu, "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions," *Journal of Research on Educational Effectiveness*, 10 (2017), 7–39.
Bailey, Martha, Brenden D. Timpe, and Shuqiao Sun, "Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency," *American Economic Review*, 111 (2021), 3963–4001.
Baker, Michael, Jonathan Gruber, and Kevin Milligan, "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, 116 (2008), 709–745.
———, "The Long-Run Impacts of a Universal Child Care Program," *American Economic Journal: Economic Policy*, 11 (2019), 1–26.
Barnett, Steve, "Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications," *Educational Evaluation and Policy Analysis*, 7 (1985), 333–342.
Barnett, W. Steven, Kenneth B. Robin, Jason T. Hustedt, and Karen L. Schulman, "The State of Preschool: 2003 State Preschool Yearbook," National Institute for Early Education Research Report, Rutgers University, 2003.
Belfield, Clive R., Milagros Nores, Steve Barnett, and Lawrence Schweinhart, "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup," *Journal of Human Resources*, 41 (2006), 162–190.
Biden, Joseph, "Fact Sheet: The American Families Plan," The White House, 2021. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/28/fact-sheet-the-american-families-plan/.
Bitler, Marianne P., Hilary W. Hoynes, and Thurston Domina, "Experimental Evidence on Distributional Effects of Head Start," NBER Working Paper no. 20434, 2014.
Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When Is TSLS Actually LATE?," NBER Working Paper no. 29709, 2022.
Bloom, Howard, and Christina Weiland, "Quantifying Variation in Head Start Effects on Young Children's Cognitive and Socio-emotional Skills Using Data from the National Head Start Impact Study," MDRC Working Paper, 2015.
Bloom, Howard S., "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8 (1984), 225–246.
Borusyak, Kirill, and Peter D. Hull, "Non-Random Exposure to Exogenous Shocks: Theory and Applications," NBER Working Paper no. 27845, 2021.
Boston Public Schools, "Introducing the Boston Public Schools: A Guide for Parents and Students," *Boston Public Schools*, 1997.
———, "Introducing the Boston Public Schools: A Guide for Parents and Students," *Boston Public Schools*, 1998.
Bowen, William G., and Derek Bok, *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions* (Princeton, NJ: Princeton University Press, 2000).
Campbell, Frances A., Elizabeth P. Pungello, Margaret Burchinal, Kirsten Kainz, Yi Pan, Barbara H. Wasik, Oscar A. Barbarin, Joseph J. Sparling, and Craig T. Ramey, "Adult Outcomes as a Function of an Early Childhood Education Program: An Abecedarian Project Follow-Up," *Developmental Psychology*, 48 (2012), 1033–1043.
Campbell, Frances A., and Craig T. Ramey, "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families," *Child Development*, 65 (1994), 684–698.
Carneiro, Pedro, and Rita Ginja, "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start," *American Economic Journal: Economic Policy*, 6 (2014), 135–173.
Cascio, Elizabeth, "Early Childhood Education in the United States: What, When, Where, Who, How, and Why," in *Routledge Handbook of the Economics of Education*, Brian McCall, ed. (London: Routledge, 2022), 30–72.

———, “Does Universal Preschool Hit the Target? Program Access and Preschool Impacts,” *Journal of Human Resources* (forthcoming).

Chaudry, A., T. Morrissey, C. Weiland, and H. Yoshikawa, *Cradle to Kindergarten: A New Plan to Combat Inequality* (New York: Russell Sage Foundation, 2017).

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR,” *Quarterly Journal of Economics*, 126 (2011), 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104 (2014a), 2593–2632.

———, “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review*, 104 (2014b), 2633–2679.

Clements, D. H., and J. Sarama, *SRA Real Math, PreK-Building Blocks* (Columbus, OH: SRA/McGraw-Hill, 2007).

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg, “Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance,” *Journal of Political Economy*, 126 (2018). https://doi.org/10.1086/699979.

Currie, Janet, “Early Childhood Education Programs,” *Journal of Economic Perspectives*, 15 (2001), 213–238.

Currie, Janet, and Duncan Thomas, “Does Head Start Make a Difference?,” *American Economic Review*, 85 (1995), 341–364.

De Haan, Monique, and Edwin Leuven, “Head Start and the Distribution of Long-Term Education and Labor Market Outcomes,” *Journal of Labor Economics*, 38 (2020), 727–766.

Deming, David, “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start,” *American Economic Journal: Applied Economics*, 1 (2009), 111–134.

———, “Better Schools, Less Crime?,” *Quarterly Journal of Economics*, 126 (2011), 2063–2115.

Duncan, Greg J., and Katherine Magnuson, “Investing in Preschool Programs,” *Journal of Economic Perspectives*, 27 (2013), 109–132.

Duncan, Greg J., and Richard J. Murnane, *Restoring Opportunity: The Crisis of Inequality and the Challenge for American Education* (Cambridge, MA: Harvard Education Press and Russell Sage Foundation, 2014).

Durkin, Kelley, Mark W. Lipsey, Dale C. Farran, and Sarah E. Wiesen, “Effects of a Statewide Pre-Kindergarten Program on Children’s Achievement and Behavior through Sixth Grade,” *Developmental Psychology* (forthcoming). http://dx.doi.org/10.1037/dev0001301.

Dynarski, Susan, Steven Hemelt, and Joshua Hyman, “The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes,” *Educational Evaluation and Policy Analysis*, 37 (2015). https://doi.org/10.3102/0162373715576078.

Elango, Sneha, Jorge Luis Garcia, James J. Heckman, and Andres Hojman, “Early Childhood Education,” in *Economics of Means-Tested Transfer Programs in the United States*, vol. 2. Robert Moffitt, ed. (Chicago: University of Chicago Press, 2016).

Felfe, Christina, and Rafael Lalive, “Does Early Child Care Affect Children’s Development?,” *Journal of Public Economics*, 159 (2018), 33–53.

Feller, Avi, Todd Grindal, Luke Miratrix, and Lindsay C. Page, “Compared to What? Variation in the Impacts of Early Childhood Education by Alternative Care Type,” *Annals of Applied Statistics*, 110 (2016), 1245–1285.

Fitzpatrick, Maria D., “Starting School at Four: The Effect of Universal Pre-kindergarten on Children’s Academic Achievement,” *BE Journal of Economic Analysis & Policy*, 8 (2008), 1897.

Friedman-Krauss, Allison H., W. Steven Barnett, Karin A. Garver, Katherine S. Hodges, G. G. Weisenfeld, and Beth Ann Gardiner, “The State of Preschool:

2019 State Preschool Yearbook," National Institute for Early Education Research Report, Rutgers University, 2019.

Garces, Eliana, Duncan Thomas, and Janet Currie, "Longer-Term Effects of Head Start," *American Economic Review*, 92 (2002), 999–1012.

Garcia, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and Maria Jose Prados, "Quantifying the Life-Cyle Benefits of an Influential Early Childhood Program," *Journal of Political Economy*, 128 (2020). https://doi.org/10.1086/705718.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor, "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica," *Science*, 344 (2014), 998–1001.

Gormley, William T., and Ted Gayer, "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program," *Journal of Human Resources*, 40 (2005), 533–558.

Gormley, William T., Ted Gayer, Deborah Phillips, and Brittany Dawson, "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology*, 41 (2005), 872–884.

Gramlich, Edward M., "Evaluation of Education Projects: The Case of the Perry Preschool Program," *Economics of Education Review*, 5 (1986), 17–24.

Gray-Lobe, Guthrie, Parag A. Pathak, and Christopher R. Walters, "Replication Data for: 'The Long-Term Effects of Universal Preschool in Boston," (2022), Harvard Dataverse, https://doi.org/10.7910/DVN/BOU7XR.

Greenberg, Erica, Grace Luetmer, Carina Chien, and Tomas Monarrez, "Who Wins the Preschool Lottery? Applicants and Application Patterns in DC Public Prekindergarten," Urban Institute, 2020.

Gupta, Snigdha, Lauren H. Supplee, Dana Suskind, and John A. List, "Failed to Scale: Embracing the Challenge of Scaling in Early Childhood," in *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*, John A. List, Dana Suskind, and Lauren H. Supplee, eds. (New York, NY: Routledge, 2021).

Havnes, Tarjei, and Magne Mogstad, "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes," *American Economic Journal: Economic Policy*, 3 (2011), 97–129.

———, "Is Universal Child Care Leveling the Playing Field?," *Journal of Public Economics*, 127 (2015), 100–114.

Heckman, James, *Giving Kids a Fair Chance* (Cambridge, MA: MIT Press, 2013).

———, "With Supports, the Impact of High-Quality Preschool Does Not Fade Out," *Washington Post February 23*, 2017.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz, "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program," *Quantitative Economics*, 1 (2010a), 1–46.

———, "The Rate of Return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 94 (2010b), 114–128.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103 (2013), 2052–2086.

Hendren, Nathan, and Benjamin Sprung-Keyser, "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 135 (2020), 1209–1318.

Imbens, Guido, and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (1994), 467–475.

Jackson, Kirabo, "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes," *Journal of Political Economy*, 126 (2018). https://doi.org/10.1086/699018.

Jan, Tracy, "Boston Preschools Falling Far Short of Goals, Study Says," *Boston Globe*, April 7, 2007.

Johnson, Rucker C., and C. Kirabo Jackson, "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending," *American Economic Journal: Economic Policy*, 11 (2019), 310–349.
Kline, Patrick, and Christopher R. Walters, "Evaluating Public Programs with Close Substitutes: The Case of Head Start," *Quarterly Journal of Economics*, 131 (2016), 1795–1848.
Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff, "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce," *Proceedings of the National Academy of Sciences*, 103 (2006), 10155–10162.
Lipsey, Mark W., Dale C. Farran, and Kelley Durkin, "Effects of the Tennessee Prekindergarten Program on Children's Achievement and Behavior through Third Grade," *Early Childhood Research Quarterly*, 45 (2018), 155–176.
Ludwig, Jens, and Douglas L. Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 122 (2007), 159–208.
Magnuson, Katherine A., Robert Kelchen, Greg J. Duncan, Holly S. Schindler, Hilary Shager, and Hirokazu Yoshikawa, "Do the Effects of Early Childhood Education Programs Differ by Gender? A Meta-Analysis," *Early Childhood Research Quarterly*, 36 (2016), 521–536.
Magnuson, Katherine A., Marcia K. Meyers, Christopher J. Ruhm, and Jane Waldfogel, "Inequality in Preschool Education and School Readiness," *American Educational Research Journal*, 41 (2004), 115–157.
Marshall, Nancy L., and Joanne Roberts, "Boston Public Schools Early Childhood Quality Study: 2010," Wellesley Centers for Women, Internal Report to the Department of Early Childhood, Boston Public Schools, 2010.
Marshall, Nancy L., Joanne Roberts, and Linda Mills, "Boston Public Schools K1 and K2 Programs Needs Assessment," Wellesley Centers for Women, Internal Report to the Department of Early Childhood, Boston Public Schools, 2006.
Martin, Naomi, "Eight Years after Walsh's Promises, Boston Prekindergarten Is Still Not Universal," *Boston Globe*, February 22, 2021.
Massachusetts Department of Elementary and Secondary Education, "Children First: A Plan for an Early Care and Education System for Massachusetts: Report of the Special Commission on Early Childhood," Massachusetts State Department of Education, Malden, 1995.
Miller, Douglas L., Na'ama Shenhav, and Michel Z. Grosz, "Selection into Identification in Fixed Effects Models, with Application to Head Start," *Journal of Human Resources* (forthcoming).
Miller, L. C., and D. Bassok, "The Effects of Universal Preschool on Grade Retention," EdPolicyWorks Working Paper Series no. 57, 2017.
Mongeau, Lillian, "Tennesse Preschool Study Receives Significant Coverage, Reaction," *Education Week*, October 9, 2015.
Muralidharan, Karthik, and Paul Niehaus, "Experimentation at Scale," *Journal of Economic Perspectives*, 31 (2017), 103–124.
National Institute for Early Education Research, "Pre-K in American Cities," National Institute for Early Education and City Health Report, Rutgers University, 2019.
Obama, Barack, "State of the Union Address," Remarks by the President, Washington DC, February 12, 2013. https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/president-barack-obamas-state-union-address-prepared-delivery.
Pages, Remy, Dylan J. Lukes, Drew H. Bailey, and Greg J. Duncan, "Elusive Longer-Run Impacts of Head Start: Replications within and across Cohorts," *Educational Evaluation and Policy Analysis*, 42 (2020), 471–492.
Petek, Nathan, and Nolan G. Pope, "The Multidimensional Impact of Teachers on Students," University of Maryland Working Paper, 2021.
Puma, Michael, Stephen Bell, Ronna Cook, and Camilla Heid, "Head Start Impact Study: Final Report," U.S. Department of Health and Human Services, Administration for Children and Families, 2010.

Puma, Michael, Stephen Bell, and Camilla Heid, “Third Grade Follow-Up to the Head Start Impact Study,” U.S. Department of Health and Human Services, 2012.
Rose, Evan K., Jonathan Schellenberg, and Yotam Shem-Tov, “The Effects of Teacher Quality on Adult Criminal Justice Contact,” NBER Working Paper no. 30274, 2022.
Rosenbaum, Paul R., and Donald B. Rubin, “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70 (1983), 41–55.
Sachs, Jason, and Christina Weiland, “Boston’s Rapid Expansion of Public School-Based Preschool: Promoting Quality, Lessons Learned,” *Young Children*, 65 (2010).
Schickendanz, J., and D. Dickinson, *Opening the World of Learning* (Iowa City, IA: Pearson, 2005).
Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores, *Lifetime Effects: The High / Scope Perry Preschool Study through Age 40* (Ypsilanti, MI: HighScope Press, 2005).
Setren, Elizabeth, “The Impact of Targeted vs. General Education Investments: Evidence from Special Education and English Language Learners in Boston Charter Schools,” *Journal of Human Resources* (forthcoming).
Thompson, Owen, “Head Start’s Long-Run Impact: Evidence from the Program’s Introduction,” *Journal of Human Resources*, 53 (2018), 1100–1139.
U.S. Department of Health and Human Services, “Head Start Program Facts, Fiscal Year 2019,” U.S. DHHS, 2019. Available at https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/no-search/hs-program-fact-sheet-2019.pdf.
Walters, Christopher R., “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start,” *American Economic Journal*, 7 (2015), 76–102.
———, “The Demand for Effective Charter Schools,” *Journal of Political Economy*, 126 (2018), 2179–2223.
Weiland, Christina, Kchersti Ulvestad, Jason Sachs, and Hirokazu Yoshikawa, “Associations Between Classroom Quality and Children’s Vocabulary and Executive Function Skills in an Urban Public Prekindergarten Program,” *Early Childhood Research Quarterly*, 28 (2013), 199–209.
Weiland, Christina, Rebecca Unterman, Anna Shapiro, Sara Staszak, Shana Rochester, and Eleanor Martin, “The Effects of Enrolling in Oversubscribed Prekindergarten Programs Through Third Grade,” *Child Development*, 91 (2019), 1401–1422.
Weiland, Christina, and Hirokazu Yoshikawa, “Impacts of a Prekindergarten Program on Children’s Mathematics, Language, Literacy, Executive Function, and Emotional Skills,” *Child Development*, 84 (2013), 2112–2130.
Whitehurst, Grover, “Rigorous Preschool Research Illuminates Policy (and Why the Heckman Equation May Not Compute),” Brookings Institution, 2017.
Willie, Charles V., and Michael J. Alves, “Controlled Choice: A New Approach to School Desegrated Education and School Improvement,” New England Desegregation Assistance Center for Equity in Education, Brown University, 1996.
Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung, “An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs,” *Journal of Policy Analysis and Management*, 27 (2008), 122–154.
Yoshikawa, Hirokazu, Christina Weiland, and Jeanne Brooks-Gunn, “When Does Preschool Matter?,” *Future of Children*, 26 (2016), 21–35.