Impacts of Oversubscribed Boston Pre-K Programs through Middle School

Christina Weiland[1], Tiffany Wu[1], Rebecca Unterman[2], Anna Shapiro[3], Shekinah Lightner[1], Thomas Staines[1], and Annie Taylor[1]

[1]University of Michigan, [2]MDRC, [3]RAND

Acknowledgements: This study was supported by grants from the Institute of Education Sciences, U.S. Department of Education, R305A220036 and R305B200011. We thank Jason Sachs, TeeAra Dias, Yuzhu Xia, the Boston Public Schools, and the Massachusetts Department of Elementary and Secondary Education for their partnership and support. We also thank Katie Hyland and Jordan Berne for outstanding research assistance and Hirokazu Yoshikawa, Allison Ryan, Douglas Clements, and Erica Greenberg for helpful feedback and insights.
Correspondence should be sent to Christina Weiland, University of Michigan, weilandc@umich.edu

Abstract

In this pre-registered study, we explored the impacts of Boston Pre-K on children's educational trajectories, school progress/engagement, and academic achievement in late elementary and middle school using lotteries for oversubscribed schools in 2007-2011 (*N*=3,092 students; 24% of all applicants). Importantly, the program was unique nationally in its strong supports for teachers and in its use of evidence-based curricula. Nearly all control-group children attended other preschool programs. Treatment compliers had markedly different educational trajectories than control compliers in grades 4-8. Treatment compliers were much more likely to be enrolled in the Boston Public Schools, less likely to be enrolled in charter schools (especially in late elementary school), and less likely to transition schools in 6th grade. Treatment compliers were 14 percentage points more likely to apply to Boston's competitive exam schools, which begin in the 7th grade ($p<.10$); they were also 10pp more likely to be admitted and 5pp more likely to enroll, though neither of these latter differences was statistically significant. These differential 6th and 7th grade transitions aligned with when larger differences by treatment status emerged on some outcomes. For example, suspension rate differences were small in 4th and 5th grades, but increased (favoring the treatment group) in the 6th and 7th grades, though only the 7th grade difference was statistically significant (i.e., treatment compliers were 7.6 percentage points less likely to be suspended in 7th grade; $p<.05$). For achievement, we found patterns favoring treatment compliers for both math and reading tests beginning in 5th grade and increasing in 6th and 7th grade, though only the 7th math effect (d=0.24 SD; $p<.05$) was statistically significant. Treatment compliers also enrolled and passed Algebra I by the end of 8th grade at a much higher rate than control compliers (22.6 percentage points, $p<.01$). We found no meaningful impacts on students' in-grade retention, special education placement, or attendance. Generalizability analyses suggested larger benefits for the full sample of BPS Pre-K applicants. Findings shine new light on longitudinal Pre-K effects in the developmentally turbulent middle school years, especially in crowded school choice environments.

Impacts of Oversubscribed Boston Pre-K Programs through Middle School

"Middle school is kind of like Middle-earth. It's a magical journey filled with elves, dwarves, hobbits, queens, kings, and a few corrupt wizards. Word to the wise: pick your traveling companions well." – Kimberley Dana, in Lucy and CeCee's How to Survive (and Thrive) in Middle School

Preschool programs have been a research focus for decades, dating back to the 1960s. The evidence on the short-term effects of these programs is clear: compared to staying home, Pre-K programs boost children's kindergarten readiness, improving their cognitive, academic, executive function, and social emotional skills (Phillips et al., 2017). Such evidence, alongside increases in maternal employment, is part of why public Pre-K programs have continued to gain ground in enrollment and in expansions. Nearly 70% of four-year-olds attend some kind of center-based preschool program in the year before kindergarten and about 43% of these are enrolled in publicly funded programs (Friedman-Krauss et al., 2024; Whitehurst & Klein, 2015).

However, how long Pre-K benefits last has been debated since the early days of Head Start in the 1960s, which represented the first major public investment in preschool in the U.S. (Cicirelli, 1969). Two consensus groups of experts reached three conclusions on the longitudinal evidence base (Phillips et al., 2017; Yoshikawa et al., 2013). First, differences in the skills of participants and non-participants shrink in the early elementary years, sometimes partially and sometimes fully. Non-participants tend to "catch up" to participants. Second, benefits of Pre-K attendance can – but don't always – re-emerge, on a wide range of behavioral, health, and educational outcomes into adulthood, even in cases where there is evidence of earlier convergence. And third, evidence from modern-day, large-scale Pre-K programs is just emerging and should be a priority for future research.

Our team represents one of the dedicated efforts in the U.S. to build this more modern-day evidence base.[1] We have partnered with the Boston Public Schools Department of Early Childhood since 2007, producing rigorous estimates of the effects of its Pre-K program on children's kindergarten readiness using a regression discontinuity design (RDD; Weiland & Yoshikawa, 2013); its impacts on children's grade retention, special education placement, and math and reading test scores through the third grade using lotteries for oversubscribed schools (Weiland et al., 2020); and descriptive studies of the program's model/features (Kabay et al., 2020; Rochester et al., 2019; Shapiro et al., 2019; Weiland et al., 2013; Weiland & Yoshikawa, 2014). Overall, similar to patterns in the broader literature, this prior work shows strong benefits on children's readiness for kindergarten and large impacts on persistent enrollment in the Boston Public Schools, but convergence of child outcomes in the K-3 years.

In the present study, we extend evidence on Boston Pre-K from 4th through 8th grade, for children who applied to oversubscribed schools from 2007-2011. Using the same lottery-based design as Weiland and colleagues (2020), we explored whether BPS Pre-K affected children's school educational trajectories, progress/engagement, and academic achievement. Following prior lottery studies (Abdulkadiroğlu et al., 2011; Weiland et al., 2020), we also examined whether findings appear to generalize to the full range of BPS Pre-K applicants in these years. In doing so, we join the small handful of studies of modern, large-scale Pre-K programs that have

[1] Other efforts include the impacts of Tulsa Pre-K and Head Start through high school (via a propensity score approach; Amadon et al., 2022); NJ's Abbott Program through high school (via a propensity score approach; Barnett & Jung, 2021); and impacts of North Carolina state Pre-K through middle school (via fixed effects leveraging differences in rollout timing; Bai et al., 2020). Another team used random assignment to estimate impacts of Head Start through third grade (Puma et al., 2012) and a fourth used propensity score and random assignment methods to estimate impacts of TN Pre-K through 6th grade (Durkin et al., 2022). Finally, multiple studies leverage school assignment algorithms (as we do) to conduct lottery-based studies of Pre-K effects. These include studies of DC's public preschool program (Baraga et al., 2024); New Haven, CT's universal Pre-K program (Humphries et al., 2024); and an older, substantially different version of the Boston program (Gray-Lobe et al., 2021).

followed children through their developmentally critical and educationally important late elementary and early adolescent years. We also make novel contributions to this knowledge base via some important outcomes new to this literature; via the unusually strong counterfactual in Boston relative to most of this literature (i.e., nearly all control group members attended some form of preschool); and via the extensive quality investments in Boston Pre-K that our team has documented (e.g., Weiland et al., 2013; Weiland et al., 2021; Weiland & Yoshikawa, 2014).

In our pre-registered exploratory study, we found that treatment compliers had markedly different educational trajectories than control compliers in grades 4-8. Treatment compliers were much more likely to be enrolled in the Boston Public Schools, less likely to be enrolled in charter schools (especially in late elementary school), and less likely to transition schools in 6th grade. Treatment compliers were 14 percentage points more likely to apply to Boston's competitive exam schools, which begin in the 7th grade ($p<.10$); they were also 10pp more likely to be admitted and 5pp more likely to enroll, though neither difference was statistically significant. These differential 6th and 7th grade transitions align with when larger differences by treatment status emerge on some outcomes. For example, suspension rate differences are small, but increase (favoring the treatment group) in the 6th and 7th grades, though only the 7th grade difference is statistically significant (i.e., treatment compliers were 7.6 percentage points less likely to be suspended in 7th grade; $p<.05$). For achievement, we found patterns favoring treatment compliers for both math and reading tests beginning in 5th grade and increasing in 6th and 7th grade, though only the 7th math effect (d=0.24 SD; $p<.05$) was statistically significant. Treatment compliers also enrolled and passed Algebra I by the end of 8th grade at a much higher rate than control compliers (22.6 percentage points, $p<.01$). We found no meaningful impacts on students' in-grade retention, special education placement, or attendance. Generalizability analyses suggest similar patterns for the full sample of BPS Pre-K applicants.

Overall, as we detail, our findings stand to inform theories and debates about the pathways through which the benefits of public Pre-K last versus fade out, including on the key question of *what kind* of Pre-K model can produce lasting impacts.

Pre-K Effects through Middle School: Theory

As reviewed in Weiland and colleagues (2020), multiple theoretical frames from economics and developmental psychology typically underpin investigations into the lasting effects (or not) of Pre-K. For example, per the human capital accumulation theory from economics, benefits may last into these years if a strong early foundation sets the stage for acquiring more advanced skills (Heckman, 2000). From developmental psychology, there are several related but distinct frames. Developmental cascades theory predicts that earlier conditions/experiences can lead to higher or lower functioning in one or more domains that then affect later competencies in multiple domains (Masten & Cicchetti, 2010). Sameroff (2009)'s transactional developmental theory posits reciprocal effects of higher early child skills on subsequent teacher behaviors, which then might have effects on child outcomes. Bailey and colleagues (2016) theorized lasting effects may occur via: 1) "sustaining environments," meaning if subsequent experiences successfully build on children's Pre-K boost; 2) a "foot-in-the-door," in which benefits may last if they get children over an important hurdle that grants them access to a beneficial program/experience or avoids a harm; and 3) skill target of the Pre-K program – i.e., for benefits to last, the skills the program focuses on must be malleable, fundamental for success, and unlikely to develop in the counterfactual.

Prior longitudinal Pre-K studies largely have not explicitly detailed how these oft-used theoretical frames fit with the specific developmental demands of the late elementary/middle school years. Said differently, in most longitudinal Pre-K studies, these theories have been applied in an imprecise way across very distinct developmental periods in the life course. In our study, similar to Phillips et al. (2018), we posit these years may hold particular importance in understanding the longitudinal effects of Pre-K and be distinct from other developmental periods. In these years, children face new internal and external demands that bring both opportunity and risk (Eccles & Wigfield, 1997). Internally, the onset of puberty brings physical and emotional changes second only to the early childhood in their reach and magnitude. These years also bring marked cognitive changes, with large increases in abstract thought and reasoning capacity (Eccles et al., 1993; Ryan et al., 2013). At the same time, due to educational policies and practices, most students experience large shifts in their schooling environments in these years. For example, students generally begin to switch classes and thus experience less opportunity for deeper connection and relationships with a primary teacher, just at a time when developmentally, they crave these kinds of relationships. Researchers have used Person-Environment fit theory to describe this misfit and its potentially detrimental consequences – i.e., this theory predicts children will thrive in environments that match their developmental needs and struggle in mis-matched environments (Eccles, 2004; Eccles & Wigfield, 1997). Empirically, the widely noted, substantial declines in academic motivation, interest in school, and achievement in adolescence begin in these developmentally turbulent years (Eccles et al., 1993; North et al., 2019; Ryan et al., 2013). Children also become more sensitive to their status in peer groups and more preoccupied with peer relations in these years (Eccles & Wigfield, 1997; Ryan et al., 2013).

Further, how children navigate these shifting internal and external demands in these years is predictive of their high school outcomes and beyond (Eccles & Wigfield, 1997).

Combining thinking from the typically used frames in longitudinal Pre-K studies with theories on adolescent development, we posit that middle school may be a particularly sensitive time for detecting Pre-K effects – i.e., that the internal/external changes children are navigating may make this period a particularly prime one for understanding patterns of Pre-K fadeout and re-emergence of benefits, especially when compared to the end of early childhood (i.e., age 8, or around the third grade). Determining what a "sustaining environment" is as children shift schooling contexts becomes more complicated than in the K-3 years. However, additional "foot in the door" opportunities arise with more accelerated coursework opportunities that typically begin in these years. And for skills, these years present new opportunities as more advanced abstract thinking is introduced, particularly in math (Clements et al., 2013).

As we review next, empirical evidence shows that lasting effects of Pre-K programs in middle childhood and early adolescence is possible but not guaranteed, especially in programs with a strong counterfactual.

Pre-K Effects through Middle School: Empirical Evidence

Empirically, studies that are diverse in methodological design, participant demographics, counterfactual, and the time period children attended Pre-K have followed children into the late elementary/middle school years and beyond. Overall, as multiple reviews have highlighted (Cascio, 2024; Phillips et al., 2017; Sawhill & Welch, 2023; Yoshikawa et al., 2013; Yoshikawa et al., 2016), Pre-K non-enrollees tend to catch up partially or fully to enrollees in the early elementary years on test scores; benefits can re-emerge in adulthood but do not always. In the late elementary and middle school period specifically, as we detail below, findings are mixed on outcomes commonly assessed in this period, such as academic test scores, school progress metrics like grade retention and special education, and behavioral/engagement outcomes. Patterns are just emerging for newer outcomes increasingly of interest to localities, such as persistence in public school system enrollment.

For academic achievement and cognitive skills, the small-scale Perry and Abecedarian randomized experiments in the 1960s and 1970s found persistent effects through middle school, in samples of Black children and compared to a counterfactual of staying home/not attending any preschool (Campbell & Ramey, 1994; Schweinhart & Weikart, 1981). In a difference-in-difference study in the early years of Georgia and Oklahoma UPK in the 1990s, reading and math test score gains in 8th grade were concentrated in children from families with low incomes, with no test score benefits for their peers (Cascio & Schanzenbach, 2013). In that study, many children from higher-income families would have attended preschool either way – i.e., the treatment-contrast was larger for children from families with low incomes who moved from no preschool to public than their counterparts who switched from private to public programs. From around the same time period (1997-2003), a randomized lottery study of an earlier version of Boston Pre-K than our focal program found no statistically significant difference of Pre-K attendance on state reading and math standardized tests, with point estimates ranging from -0.03 to 0.07 SDs across tests and years in grades 4-8 (Gray-Lobe et al., 2024). Few details about the program's quality, features, or its counterfactual are known, though supplemental evidence suggests the counterfactual was likely quite mixed, with many control group members enrolling in other preschool programs. Similarly, a recent randomized lottery-based study of New Haven UPK cohorts around that same time period found no impacts on reading and math test scores in grades 4-8 for cohorts, against a counterfactual in which virtually all control group members enrolled in other preschool programs (Humphries et al., 2024). A quasi-experimental study that took advantage of a staggered rollout of preschool funding in North Carolina found benefits on both 4th and 5th grade reading and math tests (Dodge et al., 2016). Propensity score analyses studies found benefits on math and reading in middle school in New Jersey's Abbott Pre-K program (Barnett & Jung, 2021) and in math only (and not reading) in Tulsa Pre-K (Gormley et al., 2018) and Tulsa's CAP Head Start program (Phillips et al., 2016), each against a no-preschool counterfactual. Finally, a randomized trial of Tennessee Pre-K found negative effects on reading and math test scores in 4th through 6th grade (Durkin et al., 2022), against a counterfactual in which 34% of control group children attended other preschool programs (Lipsey et al., 2018). Notably, new analysis suggests negative effects were driven by children who were induced to switch from private preschool to the TN program; there were null effects for those who in the absence of TN Pre-K would have stayed home (Bernard et al., 2024).

For special education and grade retention in K-12, a meta-analytic study of experimental and quasi-experimental evaluations published between 1960 to 2016 concluded that overall, a wide range of early education programs reduced special education placements and grade retention by about 8 percentage points (McCoy et al., 2017). The follow-up period for many of the included studies was in late elementary to middle school. Of the more recent studies not included in the review, findings are mixed on these outcomes in this developmental period. For example, impacts were null for special education in late elementary school/middle school for Tulsa Pre-K and Tulsa CAP Head Start (Gormley et al., 2018; Phillips et al., 2016) and Boston Pre-K (Gray-Lobe et al., 2024). In North Carolina, increased spending on preschool was associated with reductions in special education placement in grades 4 and 5 (Dodge et al., 2017), while in Tennessee, Pre-K participants were 3 percentage points more likely to have an IEP in 6th grade than non-participants (Durkin et al., 2022). For grade retention, effects were null in late elementary school/middle school in New Haven (Humphries et al., 2024) and in Tennessee (Durkin et al., 2022). Findings were similar in Boston Pre-K, with the exception of an increase of two percentage points in 6th grade only (p<.10). However, there was evidence of reductions in grade retention in these years in Georgia (Fitzpatrick, 2008), North Carolina (Dodge et al., 2017), Tulsa Pre-K (Gormley et al., 2018), and Tulsa CAP Head Start (Phillips et al., 2016).

For student behavior and engagement, attendance and disciplinary records are the most commonly available metrics in these years. In the age 15 follow-up, the Perry Preschool program found no benefits on nine categories of school behavior, though Pre-K attenders were less likely to be kept after class and had lower rates of self-reported delinquent behaviors (Schweinhart & Weikart, 1981). In late elementary/middle school, there were no effects on chronic absenteeism in New Haven (Humphries et al., 2024); on suspensions, truancy, absences, and juvenile incarceration in Boston Pre-K (Gray-Lobe et al., 2024); nor on absenteeism and suspensions in Tulsa Pre-K (Gormley et al., 2018). For Tulsa CAP Head Start, there was a reduction in chronic absenteeism for attenders (Phillips et al., 2016), while in Tennessee there were decreases in attendance and increases in disciplinary incidents in sixth grade for Pre-K attenders (Durkin et al., 2022). Using student surveys in Tulsa Pre-K and CAP Head Start, researchers found no differences by Pre-K attendance on achievement-related beliefs nor on sense of school belonging and engagement (Gormley et al., 2018; Phillips et al., 2018).

As highlighted in Weiland et al. (2020), the evolving Pre-K context has raised new outcomes of interest to localities that administer these programs that have less commonly been included in prior studies. These include measures of children's school contexts and educational trajectories, including persistence in the public schools, especially in more crowded choice environments. For Boston Pre-K, where an original policy goal for offering the program was attracting and retaining families in the Boston Public Schools, there were large effects of Pre-K enrollment and persistence in the public schools in the early elementary years, with 74% of treatment compliers enrolled in BPS continuously from K-3 versus 39% of control compliers (p<.001; Weiland et al., 2020). There were similar findings in the early elementary years in a lottery-based study of DC Pre-K (Braga et al., 2024). In contrast, in the study of an earlier form of Boston Pre-K, in 8th grade, 81% of control group members were enrolled in the Boston Public Schools versus 84% of Pre-K enrollees, a difference that was not statistically significant (Gray-Lobe et al., 2024). In Tulsa, Pre-K alumni were more likely to persist in the Tulsa Public Schools than were non-Tulsa Pre-K and non-Tulsa Head Start children through eighth grade (Gormley et al., 2018). Interestingly, the Boston Pre-K study we are extending (Weiland et al., 2020) found few treatment-control differences in schooling contexts in K-3, in terms of peer characteristics, average school test scores, and teacher quality proxies.

Notably, we also are adding some novel outcomes to this literature: Algebra enrollment, Algebra success, and markers of application to/acceptance to accelerated curriculum opportunities through the end of middle school. For Algebra I, we explored this outcome because taking Algebra I in 8th grade is considered a gateway course to a rigorous, college-ready curriculum, as a strong foundation in Algebra is required for any higher order mathematics in high school and completion of a rigorous mathematics curriculum in high school is correlated with post-secondary degree attainment and labor market earnings (Hein et al., 2013; National Mathematics Advisory Panel, 2008). For accelerated curriculum opportunities, this outcome has also not been widely examined in prior Pre-K studies. An exception is that in Tulsa Pre-K but not

Tulsa CAP, Pre-K children were more likely to be enrolled in honors courses in middle school; there was no difference for either program in gifted program enrollment (Gormley et al., 2018; Phillips et al., 2016).

Boston Pre-K: Model Details

As our team and BPS have extensively documented (e.g., Kabay et al., 2020; Weiland & Yoshikawa, 2013; Weiland et al., 2021), the version of the Boston Pre-K program that is the focus of this study began in 2005, when then-Mayor Menino decided Boston should offer high-quality universal Pre-K.[2] Mayor Menino argued that the program stood to better prepare kids for academic success in kindergarten and beyond and could attract families to BPS who might otherwise choose other options (Weiland et al., 2020).

As detailed in Weiland and colleagues (2020), when our focal cohorts attended the program (2007-2011), Boston Pre-K was based entirely in the public schools, paid teachers on the same scale as K-12 teachers, required teachers to have the same educational requirements of K-12 teachers (e.g., a masters degree within five years), employed at least one fulltime paraprofessional in each classroom, and was full-day school day (~6.5 hours). It was open to any child in the city regardless of income. About one-third of all four-year-olds in Boston enrolled in the program, and roughly half of all children who enrolled in BPS kindergarten had attended BPS Pre-K the year before (Shapiro et al., 2019). Class size was capped at 22.

Despite these strong structural inputs, the first-ever study of the program's quality in 2006 found that the program had significant instructional quality problems (Marshall et al.,

[2] Before then, BPS offered a smaller-scale Pre-K program that was recently the subject of a long-term follow-up study of students into adulthood (Gray-Lobe et al., 2021). Few details are available on this program's features, quality, or counterfactual.

2006).[3] Based on those descriptive findings, as explained in Weiland and colleagues (2020), the district changed their model substantially, selecting evidence-based curricula and introducing a coaching model for teachers. Specifically, the district implemented *Opening the World of Learning*, which targets children's early language and literacy skills and includes a social-skills component embedded in each unit (Schickedanz & Dickinson, 2005). For math, it implemented the *Building Blocks* curriculum, which covers both numeracy and geometry and has a heavy focus on verbal mathematical reasoning (Clements & Sarama, 2007). Importantly, these curricula have a defined scope and sequence, matching recent guidance from the National Academies on more effective curriculum choices in early childhood programs (NASEM, 2024).

Curricula implementation was supported via trainings and regular coaching, meaning weekly to bi-weekly on-site support from an experienced early childhood coach trained in both curricula. These coaches were master teachers who helped the Pre-K teachers implement the curricula and address any challenges in their classroom, such as classroom management (Weiland & Yoshikawa, 2013). Coaches also created and provided teachers with curricula implementation guides that showed how to implement the curricula together (i.e., daily/weekly schedules, how to integrate *OWL* and *Building Blocks* centers, etc.).[4]

[3] The findings landed on the front page of the Boston Globe with the headline, "Boston preschools falling far short of goals study says: Teacher quality, site safety faulted" and "Boston's public preschool and kindergarten programs are hobbled by mediocre instruction, unsanitary classrooms, and dangerous schoolyards, according to a first-ever study of the programs" (Jan, 2007 April 7).

[4] In the 2009-2010 school year, the district made changes in the supports provided to teachers and thus to the intervention model. In that year, the district provided training and bi-weekly coaching on the curricula to new Pre-K teachers only. This was because veteran teachers were already implementing the curricula with moderate-to-high fidelity (Weiland & Yoshikawa, 2013) and because the district suffered budget cuts and could not afford the same professional development model for all Pre-K teachers. Most veteran teachers thus received only training (and no coaching) from the district in 2009-2010 and 2010-2011. The exception was veteran teachers in a small number of schools (10 total, 12% of district elementary schools) that were undergoing National Association for the Education of Young Children (NAEYC) accreditation. These teachers received bi-weekly coaching on content related to the NAEYC standards, which target curricula, teacher-child and teacher-parent relationships, classrooms management, and the safety of the physical environment (NAEYC, 2008).

Boston Pre-K: Prior Evidence

Overall, Boston's model in our focal years still is fairly unique nationally, as most public Pre-K programs do not require masters degrees (and many do not require a BA); do not pay Pre-K teachers on the same scale as K-12 teachers; use more general "whole child" curriculum that lack a defined scope and sequence; and do not offer coaching to their teachers (Friedman-Krauss et al., 2024). Empirically, in our study years, Boston scored similarly to other large-scale public preschool programs in the U.S. on standardized observational measures of its structural, organizational, and emotional support quality but considerably higher on instructional quality (Chaudry et al., 2021; Weiland & Guerrero Rosada, 2022; Weiland et al., 2013). The program had the highest observational instructional quality scores of any large-scale Pre-K program with such data at the time, scoring for example, 2.4 SDs higher than recent Head Start quality estimates nationally (Weiland, 2016).

An age-cutoff-based regression discontinuity study showed moderate-to-large benefits on the language, literacy, mathematics, and executive function skills at kindergarten entry of children who attended the program in 2008-2009 (Weiland & Yoshikawa, 2013). It also had small impacts on developmentally important social emotional and executive function outcomes that were not as targeted by the curricula. These effects were achieved against a counterfactual in which about two-thirds of comparison group children experienced some form of non-parental care in the prior year (and 57% overall were in other preschool programs; Weiland & Yoshikawa, 2013).

Unfortunately, the age-cutoff-based regression discontinuity design cannot be used to estimate longitudinal impacts as the comparison group gains access to preschool in the year after the treatment group, once they are age-eligible. Accordingly, Weiland and colleagues (2020) used a lottery-based strategy to estimate causal impacts of the Boston Pre-K program in early elementary school. In Boston, as in many other large cities around the country (Abdulkadiroğlu et al., 2011; Baraga et al., 2024; Bloom & Unterman, 2014), children are assigned to schools using the deferred-acceptance algorithm that creates lotteries for oversubscribed schools. Leveraging these lotteries, Weiland and colleagues (2020) found no differences by Pre-K enrollment in retention in grades K-2, special education placement in K-3, and standardized reading and math test scores in third grade. However, there was a large effect on Boston Pre-K attenders' enrollment and persistence in the Boston Public Schools from K-3. Additional work showed: (1) that K-3 impacts varied substantially across schools; and (2) the Pre-K boost persisted through third grade for children who applied to and attended elementary schools with stronger third-grade test scores and a better track record in improving students' scores (Unterman & Weiland, 2023).

Important for gauging external validity, the lottery sample (children who applied to oversubscribed programs) represented only about a quarter of all applicants and were somewhat more advantaged than the full applicant population. Further, virtually all lottery control group children attended other preschool programs (88% of control compliers). In economics terms, as Cascio (2021) described the prior study: "finding of null effects reflects essentially complete formal program substitution on the margin" (p. 42). In a quasi-experimental analysis using a propensity score weighting approach validated on the lottery sample, there was a pattern of small persistent benefits on those same outcomes for the full sample of children who enrolled in the program in between 2007 and 2011 and, for those enrolled in BPS in K-3, on children's K-3 literacy test scores (Weiland et al., 2019; Weiland et al., 2020). The present study uses the same lotteries and 2007-2011 sample to extend evidence on Boston Pre-K through 8th grade.

Current Study

Following Weiland and colleagues (2020), we used data from four cohorts of students who applied to the Boston Public Schools (BPS) Pre-K program between the 2007-2008 and the 2010-2011 school year to investigate the effects of enrolling in the Boston Pre-K program versus students' other options. Specifically, we leveraged *oversubscribed first-choice lotteries* to address our research question: What is the impact of Boston Pre-K on children's educational trajectories, school progress/engagement, and academic achievement in late elementary and middle school?

Method

Data set. We build on Weiland and colleagues (2020)'s study of Boston Pre-K through third grade. The original dataset was constructed using data on students' choices and baseline demographics during the BPS assignment process from the spring of the 2006–2007 through 2009–2010 school years (for enrollment in 2007– 2008 through 2010–2011). In this new pre-registered analysis, using each student's unique identifier, we merge on district and state administrative records to follow this same sample from fourth grade through the end of eighth grade.[5]

Sample. Our sample consists of children whose families applied for a BPS Pre-K slot between 2007 and 2010 and were randomized to an offer of enrollment in the program or not. This is the same sample as in Weiland and colleagues (2020). In all, 12,740 children applied in

[5] See the Registry of Efficacy and Effectiveness Studies, 13340.1v1. Note that we pre-registered analysis through early adulthood. COVID-19 disruptions to students' high school years, in different grades, are complex. Missing data patterns in high school vary by cohort and cohorts also faced different policy requirements. According, these findings will appear in future papers.

our study years and ultimately, 3,092 were in a randomized lottery for their first-choice school.[6] In Appendix B Table 1, we show the distribution of applicants by year, along with the percent of BPS elementary schools with a Pre-K program represented in the lotteries.

Importantly, as explained in Weiland and colleagues (2020), the lottery process resulted in treatment and control groups that were balanced in terms of their background characteristics. As we show in Table 1, there were 2 (out of 12) statistically significant differences between the two groups – lottery winners were 5.7 percentage points less likely to be Latino (*p* <0.01) and 2.9 percentage points more likely to be Asian (*p* <0.01). A joint F-test used to assess the statistical significance of the overall difference between the first-choice lottery winners and control group members could not reject the null hypothesis that there was no difference between the two groups (*p* =.213).

Demographically, our lottery sample was diverse in their background characteristics (see Table 1). For example, 35% of lottery winners and 41% of control group members identified as Hispanic; about a quarter overall as Black; another quarter as White; and the rest (~13%) as Asian, multi-racial, or other. About 57% of lottery winners were eligible for free/reduced lunch (56% for control group children). A little over half spoke English at home, a quarter spoke Spanish, and ~20% spoke another home language. In Table 1, we also display demographics characteristics of all Boston Pre-K applicants. Overall, lottery sample members were

[6] Weiland and colleagues (2020) reported 3,182 lottery participants. Upon continued analysis of the BPS assignment data, the team found 90 students were inaccurately included in the analytic sample due to a coding error as they were part of lotteries missing either a treatment or control group member and thus needed to be dropped. All analyses in this paper have excluded these students; updated main impact findings from Weiland and colleagues are presented in Appendix B Table 9. The new sample's ITT impacts generally fall within a tenth of a point of the prior reported results, the largest difference that enrolled in BPS from K-3 and ever enrolled in BPS are two percentage points lower for the new sample, though still statistically significant.

proportionately more White and less likely to eligible for free/reduced/lunch than the full sample of applicants. We return to these differences and to generalizability in the Results section.

Lottery assignment details. In our study years, under the BPS school choice plan, in the winter and spring, families ranked up to 10 schools that offered Pre-K that they wanted their children to attend the following fall. For parsimony, see Weiland et al. (2020) for full details and history on this process. In brief, unlike in most other school choice programs, in BPS, there are no automatic neighborhood schools. Families have different a priori priorities to different schools based on criteria set by the district that amount to different combinations of sibling and geographic (walk zone) priorities. There were four different deadlines by which to submit rankings and generally, families who submitted their rankings earlier were more likely to get one of their choice schools. Families could re-apply to later rounds (e.g., if they did not like their assigned choice school from early rounds or if they relocated to a different attendance zone). Families were assigned different priorities to different schools based on criteria set by the district, such as sibling and walk zone priority, sibling priority only, walk zone priority only, and no priority (listed in order from most to least priority). Key to our design, there was more demand for Pre-K than slots in 67-83% of BPS schools with Pre-K programs in our study years. In such cases, the assignment algorithm used a random number (unknown to the families at all stages of the process) to break ties between students with the same priority to a given school. From students' applications, we identified first-choice lotteries among students with the same preference to the same oversubscribed school/program (Abdulkadiroğlu et al., 2011; Bloom & Unterman, 2014; Weiland et al., 2020).

As in Weiland and colleagues (2020), 91% of lottery winners who were offered the opportunity to enroll in the BPS Pre-K program did so, with about 90% of lottery winners enrolling in their first-choice school, 2% enrolling in a school not in their initial choice list, and 9% not enrolling in Boston Pre-K. By definition, all control group students lost their first-choice lottery but ultimately, 62 percent enrolled in Boston Pre-K by winning a seat in one of their lower-ranked schools, coming off of a waitlist, or participating and being assigned in a subsequent school assignment round.[7] The compliance rate accordingly was ~29 percentage points (91% minus 62%). This difference is on the lower end for randomized experiments but similar to other studies utilizing naturally occurring lotteries within choice processes (Abdulkadiroglu et al., 2015; Angrist et al., 2016). We leveraged these lotteries (as described later in this section) to estimate the effect of *enrolling* in Boston Pre-K for compliers, defined as (1) treatment group members who won an over-subscribed first-choice lottery and enrolled in Boston Pre-K and as (2) control group members who lost their first-choice lottery and did not enroll in any Boston Pre-K.

Counterfactual. Prior preschool studies show that the counterfactual matters greatly in preschool impact studies (Feller et al., 2016). Using district and state administrative records from the 2007-2008 and 2008-2009 cohorts,[8] Weiland and colleagues (2020) found that: (1) nearly all (97%) of control group members enrolled in some form of preschool; and (2) 88% of lottery control group children who complied with their random assignment and did not enroll in Boston Pre-K were enrolled in another center-based program. In comparison, in other recent preschool evaluations, up to about half the control group has attended other center-based preschools program (e.g., likely 0% in New Jersey; 34% in Tennessee, about 50% in Head Start; Barnett &

[7] For this group, Weiland and colleagues (2020) reported that roughly 13% of the control group enrolled in their first-choice school, 29% enrolled in a school lower on their choice list, 23% enrolled in a school not on their initial choice list, and 35% did not enroll in Boston prekindergarten.

[8] As explained fully in Weiland and colleagues (2020), data availability limited counterfactual analysis to the first two of the four cohorts of interest in that prior study and in our present study.

Jung, 2021; Bloom & Weiland, 2015; Lipsey et al., 2018). The exception is New Haven UPK, where virtually all children who lost a lottery for a UPK seat enrolled in another preschool program (Humphries et al., 2024). The counterfactual in Weiland and colleagues (2020) is the same as in our extension of that study – i.e., Boston Pre-K as compared to a particularly strong counterfactual.

Measures

We created our study measures using administrative data from the Massachusetts Department of Elementary and Secondary Education and the Boston Public Schools. We conceptualized our outcomes measures in three categories – educational trajectory, school progress/engagement, and academic achievement. Due to mixed evidence on Pre-K effects in late elementary and middle school, we pre-registered these outcomes as exploratory with the Society for Research in Educational Effectiveness registry. We did not adjust for multiple hypothesis testing, following recommended guidance for exploratory studies (Schochet, 2009). Except where noted, we constructed outcomes by follow-up year and not the grade a given student was in, to account for grade retention patterns. However, for interpretation, we generally refer to grade levels following student's expected and actual grade progression (i.e., in follow-up year 9, most students were in 8th grade). Because our study is an upward extension of Weiland et al. (2020) which followed children through third grade, we focus on children's 4th-8th grade outcomes.

Educational trajectory outcomes. We examined whether BPS Pre-K affected children's subsequent educational trajectories, using proxies for BPS enrollment and persistence, enrollment in charter schools, enrollment in competitive accelerated programming, application and enrollment in BPS's exam-based middle schools, and subsequent schooling context.

For BPS enrollment and persistence, we set the enrollment variable for each follow-up year to 1 if the student enrolled in BPS at least one day and to zero otherwise. We also constructed measures of enrollment in BPS across grades 4-5 (elementary school years) and 6-8 (middle school years). We followed a similar procedure to create charter school enrollment variables.

We also examined several outcomes related to BPS's Advanced Work Course program, a program the district has operated since before school desegregation in 1974 (BPS, 2013; Cohodes, 2020). AWC offers an accelerated curriculum to academically advanced students in reading and math. Students are offered slots in the program in 4th and 5th grades based on their scores on a standardized test used only for AWC placement in the district and in 6th grade, on the basis of their 5th grade test scores (and with alternative exams offered for Spanish-speaking ELL students). A large-scale regression discontinuity study of AWC in cohorts in 3rd grade in 2001-2005 found AWC had positive impacts on students' test scores and improved their high school graduation and college enrollment, with gains driven by Black and Latino students (Cohodes, 2020). Importantly, all students in BPS took the exam for AWC entry in our study years, students outside BPS are allowed to take the test as well, and around the top 15% of test takers are offered admission. To accept admission, families must complete a school choice form and some students then must change schools, if the program is not offered in their elementary school. Because of this feature of the program, we created a measure of whether the student was enrolled in a BPS school that offered AWC in 3rd grade (i.e., the year before the first grade in which they could enroll in AWC), as well as a measure of whether each student was ever enrolled in AWC. Ideally, we also would have constructed a measure of whether students took the AWC test and whether they were accepted into the program, but BPS did not share these data with our team.

We also created dichotomous measures of whether students ever took the optional exam school test in 6th-8th grade, were ever offered admission to an exam school, and whether they enrolled. Admission in our study years was based on a combination of grades and test scores on the Independent School Entrance Exam (with grades and test scores equally weighted). The entrance exam was optional for students and was offered at no cost on a Saturday in early November each year. Any student who lives in the city of Boston can apply (they do not have to attend BPS). An analysis covering 2007-2013 found about 35% of eligible students took the exam, 17% were admitted to an exam school, and 15% enrolled (Goodman & Rucinski, 2018). Of those offered admission, 87% enrolled. Prior work found no benefit of just being admitted to these schools vs. just missing the cutoff on state standardized test scores, PSAT and SAT scores, number of AP tests, college attendance, or college selectivity (Abdulkadiroğlu et al., 2014). However, these are effects for the marginal student, not all students. Exam schools are very sought-after and may be key to understanding whether/how Pre-K attendance influences students' trajectories.

Finally, we used publicly available school-level data from the Massachusetts Department of Elementary and Secondary Education to describe the characteristics of the schools students attended in each year from kindergarten through 8th grade. We merged these school-level variables on to each student record in each school year, matching the student with the characteristics of the school they attended for the longest period of time in a given year. To describe student sociodemographic characteristics, we used measures of the percentage of students from low-income or high-needs families, the percentage of English learners, the percentage of students who speak a language other than English at home, the percentage of students with disabilities, the percentage of students who were African American, Asian, Hispanic, or White, and the percentage of students who were female. To describe the academic context, we used the percentage of students who were proficient or advanced on the state English Language Arts and mathematics tests in grades 3-8. Finally, we also used measures of the percentage of licensed teachers, the percentage of teachers who were rated as proficient and the percentage of teachers rated as exemplary in the state's teacher rating system (available from 2009-2010 onward), the percentage of teachers still teaching in a school from the previous school year (stability rate of teachers), the percentage of students who remain in their school for the full school-year (stability rate of students), the student-teacher ratio, and the average class size. To construct our analytic variables, we created composite measures for grades K-5 and for grades 6-8 to describe students' elementary and middle school contexts. For these composite measures, we averaged each characteristic across all available years. If data were missing for a student in a given year (e.g., first grade), we used non-missing data to compute the composite measure (e.g., K, second, third, fourth, fifth).

School progress/engagement outcomes. From administrative records, we constructed year-by-year measures of children's in-grade retention, special education status (i.e., having an Individualized Education Plan), attendance, and discipline across their 4^{th}-8^{th} grade years. We also constructed composite measures for each measure across these grades (i.e., ever retained in 4-7). For attendance, we calculated student attendance rate by dividing each student's number of absence days by the total days the student was enrolled in their majority-enrolled school during a given year and subtracting this from 100. Using a rate instead of a raw count of days attended accounts for variation in enrollment length, both between public and charter schools and for students who transfer schools mid-year. In our robustness checks, findings are not meaningfully different from other plausible attendance measures sometimes used in the literature, such as the number of days absent (e.g., Gottfried, 2009; Gray-Lobe et al., 2021). Finally, we created measures of whether a student was suspended in each year 4th-8th and whether they were ever suspended in any of these grades.

Academic achievement. We used children's Massachusetts state standardized tests in reading and math in 4th through 8th grades in our focal cohorts. Across cohorts, the state used three different tests in our study years. Up until 2014, students were administered the Massachusetts Comprehensive Assessment System (MCAS), now called "Legacy MCAS." In 2015 and 2016, districts were able to choose to administer either Legacy MCAS or the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment, an exam based on Common Core standards. In 2015, 54% of districts in Massachusetts administered PARCC, and in 2016, 67% of districts did so. In 2015 and 2016, in the three largest school districts in the state—Boston, Worcester, and Springfield—individual schools chose which test to administer (MA DESE, 2016). In 2017, all districts administered a new version of MCAS ("Next-Gen MCAS"). We followed state guidance on how to pool these tests across years (MA DESE, 2022). For Legacy MCAS through 2014, we use the students' raw scores converted into z-scores. For PARCC and Legacy MCAS in 2015 and 2016, we use state-provided thetas. For Next-Gen MCAS in 2017 and onwards, we use state-provided thetas. In all cases, we standardize to Boston Public School students in the requisite grade (by year) so that means can be interpreted in comparison to the average BPS student in that grade. Note that our final cohort was in 8th grade when the pandemic began and test scores are not available for them (an issue that affected this outcome/cohort only). Thus, we have 8th grade test scores for three out of our four cohorts.

Finally, we also created indicators of taking Algebra I in 8^{th} grade or earlier and earning a passing grade in Algebra I in 8^{th} grade or earlier.

Covariates. Finally, we used the same set of student-level covariates as in Weiland et al. (2020). For race/ethnicity, we used dichotomous indicators for whether a student was Asian, Black, Hispanic, White, or mixed/other. We also used dichotomous indicators that identified whether students' home language was English only, Spanish, or another language; whether each student was eligible for free-reduced priced lunch; whether the student was male; and whether the student's country of origin was the U.S.

Data Analytic Plan

We leveraged first-choice lotteries for BPS Pre-K seats to estimate intent-to-treat (ITT) effects (Abdulkadiroğlu et al., 2011; Bloom & Unterman, 2014) and complier average causal effect (CACE; Gennetian et al., 2005). We essentially extended the approach used in Weiland and colleagues (2020) to additional outcomes. As explained in that prior work, we chose to constrain our sample to first-choice lotteries because "when a student is competing in any lottery other than her first lottery, her probability of being assigned to a lower choice may depend in part on her earlier choices (and not just her random number) and thus using these later lotteries could pose a threat to randomization..." (Weiland et al., 2020, p. 1408). Focusing only on students' first-choice lotteries ensures that we have identified a purely experimental sample. Within this sample, random assignment assures that the treatment and control groups were, in expectation, equivalent in all measurable and unmeasurable characteristics. We estimate, for each lottery, differences in mean outcomes for winners and control group members, and then average the results across lotteries.

Our ITT model was the same as in Weiland and colleagues (2020):

$$Y_{ij} = \beta T_{ij} + \sum_{k=1}^{K} \pi_k I_{kij} + \sum_{p=1}^{P} \theta_p X_{pij} + \varepsilon_{ij} \quad (1),$$

where Y_{ij} is the relevant outcome for student *i* in lottery *j*; T_{ij} is a lottery winner indicator equal to 1 if student *i* wins lottery *j* and 0 otherwise; I_{ij} is a set of *k* lottery indicators equal to 1 for lottery *j* and 0 otherwise; X_{ij} is a set of p student-level covariates (race/ethnicity, gender, eligibility for free or reduced-price lunch, age, country of origin and home language status); and ε_{ij} is a random error for student *i* that is clustered by the Pre-K school that students entered after their lottery. For students that who did not enroll in the program – the majority of whom are in the control group – we assume they are not clustered together in another setting. In equation (1), the parameter of interest is the $\hat{\beta}$ coefficient, which identifies the effect of winning a lottery on student outcomes. Its associated t-statistic identifies statistical significance.

We next use students' first lottery participation as an instrument for estimating the complier average causal effect (CACE; Gennetian et al., 2005), or the effect of enrolling in Boston Pre-K. Per above, compliers are the subgroup of treatment students who won their first-choice lottery and enrolled in BPS Pre-K, compared with control group members who lost their first-choice lottery and ultimately did not enroll in BPS Pre-K. As in Weiland and colleagues (2020), because the overwhelming majority of lottery winners enrolled in their first-choice school, our enrollment effect represents *the effect of enrolling in children's first-choice program*, versus not at all.

To estimate CACE, we used the same two-stage equations as in Weiland and colleagues (2020). This approach uses a single-instrument model (i.e., a Wald estimate) to avoid finite sample bias from "weak instruments" (Bound et al., 1995). This approach is well suited to our work especially because some of our first-choice lottery blocks are quite small and has also been used in other randomized experiments and lottery-based studies (Abdulkadiroğlu et al., 2011;

Bloom & Unterman, 2014; Gennetian et al., 2005; Ludwig & Kling, 2007). Each students' lottery result is a valid instrument because it was randomized and thus cannot be correlated with error terms for student outcomes (i.e., no third path; Bloom et al., 2010). Our first stage was:

$$E_{ij} = \beta T_{ij} + \sum_{k=1}^{K} \pi_k I_{kij} + \sum_{p=1}^{P} \theta_p X_{pij} + w_{ij} \quad (2),$$

where E_{ij} is an enrollment indicator equal to 1 if student *i* ever enrolled in BPS Pre-K and 0 otherwise, and all other terms are defined as in equation (1). Our first-stage F-statistic equals 11, which is *just* above the recommended threshold for instrument strength (Bloom et al., 2010; Bound et al., 1995).

We then specified our second stage equation as:

$$Y_{ij} = \delta \hat{E}_{ij} + \sum_{k=1}^{K} \pi_k I_{kij} + \sum_{p=1}^{P} \theta_p X_{pij} + e_{ij} \quad (3),$$

where $\hat{E}_{ij}$ equals the fitted value of the enrollment outcome from the first-stage equation, e_{ij} is a random error that is clustered by the Pre-K school that students entered after their lottery, and all other terms are defined as in equation (1). Our parameter of interest in equation (3) is δ, which represents the average effect of enrolling in BPS Pre-K for compliers. Per the What Works Clearinghouse, our ITT estimates meet the highest standard of evidence, while our CACE approach is considered quasi-experimental (What Works Clearinghouse, 2014).

Finally, there was a small amount of missing data on all covariates except age in our lottery sample, ranging from 0.4% to 3.6% (see the Table 1 note). Because prior work with this sample found no sensitivity of findings to missing data, we used a complete case approach in the present study (see Robustness Check section; findings are robust to multiple imputation).

Results

Attrition analysis. Overall, we found little evidence that attrition was a threat to internal validity, though data were missing at higher rates in the treatment than in the control group. In

Appendix A, we provide missingness rates for all our outcomes in the treatment versus control groups, along with results of tests of the differences in these rates. Educational trajectory outcomes missingness rates ranged from 0-26%, with differential missingness by treatment status of 0-9 percentage points that is not statistically significant (with the exception of percent teachers rated exemplary; Appendix A Table 1). For school progress/engagement outcomes, missingness rates across outcomes ranged from 3-23% (Appendix A Table 2), with differential missingness rates of up to 9 percentage points that were not statistically significant. In Appendix A Table 3, we show missingness for test scores and Algebra outcomes, with missingness rates from 15-25 percent and differential missingness by treatment status of 3 to 8 percentage points that is not statistically significant. The exception was 8th grade test scores (missing for 43 percent of treatment students and 46 percent of control students; as we explained, state standardized tests were cancelled when most of our last cohort was in 8th grade).

We also examined baseline balance checks for non-attritors, on three different samples – children with non-missing reading or math state standardized scores in 7th grade, children with any non-missing test score data in 4th through 8th grade, and children with non-missing data on taking Algebra by the end of 8th grade. As shown in Appendix A Tables 4-6, results were very similar to balance checks for the full randomized sample, with joint *F* tests that were not statistically significant.

Educational trajectories. In Figure 1, we display a visual representation of our sample's educational trajectories through 8th grade. Overall, the figure and our ITT and CACE estimates in Table 2 show markedly different educational trajectories by treatment status in the focal period of this study, in 4th-8th grade.

For BPS enrollment, winning a first-choice Pre-K lottery caused statistically significant increases in the likelihood of enrolling in BPS each grade from 4th-8th grades of about 5-9 percentage points (ITT effects). For compliers (CACE), enrolling in Boston Pre-K led to large, statistically significant increases in 4th-8th grade enrollment in BPS, ranging from 19-31 percentage points across these grades. Complier means show that 60% of treatment compliers were in BPS in grades 4 and 5, versus 30% of control compliers. For the middle school years, 44% of treatment compliers were persistently in BPS in grades 6-8 versus just 27% of control compliers.

Treatment children were also less likely to enroll in a charter school in these years than control compliers, though differences were statistically significant only in 4th and 5th grades (Table 2). Treatment compliers were 6 to 14 percentage points less likely to enroll in a charter school across grades. Interestingly, the largest jump in treatment complier enrollment in charters occurred between 4th and 5th grades; in 4th grade, 5% of treatment compliers were in charters and in 5th grade, 16%. Control compliers show the largest jump at this same transition point, though the magnitude is smaller, with 19% in charters in 4th and 25% in 5th.

For Advanced Work Course (AWC), as shown in Table 2, control compliers were enrolled in AWC-offering BPS schools in 3rd grade at higher rates than treatment compliers. Specifically, 32% of control compliers were in AWC-offering BPS schools in 3rd grade, versus 20% of treatment compliers, a difference of 12 percentage points (p<.10). Recall that to accept AWC admission, families must complete a school choice form and some students then must change schools, if the program was not offered in their elementary school. Accordingly, treatment compliers faced higher administrative and transitional burden to enroll in an AWC. Likely reflecting these higher burdens, for AWC enrollment from 4th-6th, we found a 12-percentage point difference favoring the control group (*p*<.05). Descriptively, conditioning on enrollment in an AWC-offering school in 3rd grade, rates of ever enrolling in AWC was 36% in the treatment group versus 40% in the control group. Since we lack information on AWC offers, we interpret our AWC findings as further evidence that treatment and control children on average were on different educational trajectories after Pre-K.

We find further evidence of different trajectories by treatment status when turning to Boston exam schools, which offer an accelerated curriculum starting in 7th grade. Pre-K led to a marginally significant increase in applying to exam school in middle school of about 4pp (ITT) and 14pp for compliers (*p*<.10). The CACE effect on being admitted to an exam school in middle school was 10pp and of enrolling in an exam school, 5 pp, both favoring treatment compliers, though neither was statistically significant.[9] In an extension analysis, we created a school-level measure of the percent of sixth graders that took the exam school test in each year – i.e., a proxy for exam school test taking culture – and fit impact models with this measure as the outcome. The ITT was 5.48pp (*p*<.001); the CACE was 18.58pp (*p*<.001), with a treatment complier mean of 35 percent versus a control complier mean of 16 percent. These findings suggest part of the exam school mechanism was that treatment complier students were in schools

[9] Conditional on taking the test, 65% of treatment compliers were admitted to exam schools, versus 59% of control compliers. This analysis is endogeneous but informative; the larger share of treatment compliers admitted is not solely due to a large share taking the exam school test. In addition, restricting to children who had an enrollment record in the state of Massachusetts in the follow-up year, ITT and CACE effects are larger for taking exam school test, being admitted to an exam school, and/or enrolling in an exam school – i.e., nearly half of treatment applicants applied to exam schools versus just over a quarter for control compliers (a 21 pp difference; *p*<.01). Treatment compliers were 16 percentage points more likely to be admitted (*p*<.10) and 13 percentage points more likely to enroll in exam schools than control compliers, though this latter difference was not statistically significant.

in 6th grade in which taking the exam school test was much more common than were control complier students.

In another extension analysis, we also find that treatment students are less likely to experience a school transition between 5th and 6th grade than are control students. As shown in Appendix B Figure 1, treatment compliers were 24 percentage points less likely to be in the youngest grade offered in their school in 6th grade (and correspondingly 19pp less likely to be in the oldest grade offered in their school in 5th grade). As described above, we find an increase in BPS persistence for treatment students, which offers K-8 schools; an increase in exam school enrollment, which begins in 7th grade; and a decrease in charter school enrollment, many of which begin in 5th grade. The net effect of these differential trajectories is reflected in the observed decrease in the likelihood they experience a school transition between 5th and 6th grade.

In Table 3, we display CACE impacts of Boston Pre-K on children's K-5 and 6-8 educational contexts – i.e., the characteristics of their schools most attended in their elementary and middle school years (see Appendix B Table 2 for ITT results). Overall, K-5 contexts differed little by treatment status, with a few exceptions. Treatment compliers attended schools that enrolled proportionately fewer Black and Asian students (9pp and 3pp, respectively, $p<.05$) and more Hispanic students (10pp, $p<.10$). Treatment compliers enrolled in schools with more licensed teachers (4pp, $p<0.05$), more teachers rated as exemplary on teacher evaluations (4pp, $p<0.05$), and higher rates of student enrollment stability (2pp, $p<0.05$). In middle school (Table 3), there were some shifts in these contextual schooling characteristics, though again, contexts were more similar than different by treatment status. The peer mix differed by treatment status, with proportionally more students identified as English learners (4.6pp, $p<0.05$) and speaking a language other than English at home (5.7pp, $p<0.05$), proportionally more Hispanic peers

(10.6pp, *p*<.01), and proportionally fewer White peers (-7.0 pp, *p*<.05). For test scores, treatment compliers enrolled in schools with fewer children scoring proficient/advanced on state math test scores (6pp, *p*<0.10).

In sum, treatment and control children were on markedly different educational trajectories in 4th-8th grades in terms of school types and access to accelerated curriculum programs. However, their schooling environments were qualitatively more similar than different in terms of peer characteristics, teacher/school characteristics, and school-level test scores.

School progress/engagement. As we show in Table 4, impacts for in-grade retention were null across grades, with ITT and CACE coefficients that were close to zero in magnitude for each grade from 3rd-7th and on an ever retained 3-7 composite. Complier means align with overall BPS trends in this age span. In 2018, the percent retained in Boston in grades 4-8 ranged from 0.5% to 1.4% across grades (MA DESE, 2024). Nationally, the yearly annual retention rate was 1.5% in 2010 (Warren et al., 2014).

For special education, ITT and CACE estimates were also null across grades. For ever placed in special education from 4th-8th, the ITT estimate was 1.33 percentage points and for CACE, 4.48, though neither was statistically significant. For context, Massachusetts has one of the highest IEP rates in the country, at 20% versus the national average of 15% (National Center for Education Statistics, 2024). Approximately 20.3% of BPS elementary-school students in 2018-2019 had an IEP (MA DESE, 2025). As shown in Table 4, our complier means are consistent with these rates.

For attendance (Table 4), magnitudes were small and close to zero on our 4th-8th composite and ranged across grades from –0.07 to –0.48 percentage points (ITT) and –0.25 to -

1.66 percentage points (CACE).[10] The impacts in grades 5-7 were statistically significant, with the treatment group missing more school. However, the magnitudes of these differences were small and their statistical significance is not robust across plausible alternative constructs commonly found in the literature (see Robustness Checks and Appendix B Table 6). Across these grades, treatment and control compliers both had very high attendance, between 95 and 97 percent. In the 2018-2019 school year, BPS average attendance was 92% (MA DESE, 2019).

Also in Table 4, student suspension ITT and CACE differences are modest and not statistically significant, with the exceptions of 6th and 7th grade. In 6th grade, 6.3% of treatment compliers were suspended versus 9.0% of control compliers, though this difference is not statistically significant. In 7th grade, 6.5% treatment compliers were suspended versus 14.1% of control compliers, a statistically significant difference of 7.6pp (p<.05). On our summary measure across 4th-8th grades, treatment compliers were 5.5pp less likely to be suspended, though this difference was not statistically significant.

Academic achievement. Finally, in Table 5, we display impacts on children's state standardized test scores and Algebra-related outcomes. For reading, impacts across grades were not statistically significant but favored the treatment group (i.e., 0.02-0.05 SDs for ITT; 0.05-0.17 SDs for CACE). For math, impacts ranged from -0.00 to 0.07 SDs for ITT and for CACE, -0.00 to 0.24 SDs. The impact for math was statistically significant in 7th grade only (p<0.05).

[10] Due to disruptions in attendance and enrollment data collection following the transition to virtual learning during the pandemic in March of 8th grade for our final cohort, we refit attendance impact models excluding this cohort; the resulting estimates for 8th grade attendance (ITT = 0.08, p = 0.80) and average attendance across grades 4–8 (ITT = –0.00, p = 0.99) were not substantively different from the main results including all four cohorts.

Complier means highlight that the treatment group scored 0.37-0.46 SDs across grades higher than the average BPS student; control compliers scored 0.20-0.41 SDs higher.

For Algebra, we found statistically significant positive impacts on Algebra enrollment by end of 8th grade and on earning a passing grade in Algebra. ITT effects were about 6.6 percentage points for Algebra enrollment (p<.01) and 6.5pp for earning a passing in Algebra I *(p*<.05). For CACE, Boston Pre-K enrollment led to an increase of 23pp in Algebra enrollment *(p*<.01)*,* with a complier treatment mean of 69 percent versus 46 percent for the control group. For passing Algebra, complier means were 62 percent for the treatment group versus 39 percent for the control group, meaning a much larger share of the treatment group had not only enrolled in Algebra but also was on track for more advanced math coursework in high school than in the control group.

Notably, Algebra impacts also appear convergent with standardized test impacts – i.e., in magnitude, impacts were strongest on test scores in the 7th grade, when presumably decisions about math course taking in 8th grade was being decided. This is also the school year when exam schools start in Boston, where we found impact differences favoring the control group. We also found suspension rates differences favoring the treatment group in grade 7 as well. Descriptively, of students enrolled in Algebra I or higher by the end of 8th grade in our sample, 42% were in a BPS exam school, 40% were in another BPS school, 11% were in charter schools, and 13% were in other Massachusetts public school districts. In an extension analysis, we created measures of the percent of 8th graders taking Algebra in each sample member's 8th grade school. We found that treatment compliers were in schools in which 8th grade students overall were substantially more likely – 19 percentage points *(p*<.01) – to be enrolled in Algebra I (or higher) than control compliers. Overall, it appears treatment compliers were in substantially more ambitious math instructional school environments in 8th grade and were also scoring higher on math in the years leading up to 8th grade.

Robustness checks. We also examined the sensitivity of our findings to several plausible alternative modeling decisions. First, we show that school context findings were similar in 5th grade (Appendix B Table 3) and 7th grades (Appendix B Table 4), versus our primary approach of creating K-5 and 6-8 averages. Next, we examined sensitivity of our Algebra findings to defining variables in terms of the end of eighth grade (our primary approach) versus by the end of follow-up year 9. As we show in Appendix B Table 5, findings are not meaningfully different across approaches, with slightly larger magnitudes with the alternative approach. In a similar spirit, we show in Appendix B Table 6 that our attendance estimates are similar in their small magnitude across other plausible attendance constructs (i.e., number of days attended and number of unexcused absences) and that statistical significance is not robust. We also refit impact models in which we adjusted for missing data on covariates via multiple imputation with 50 datasets (instead of our primary strategy of listwise deletion). As shown in Appendix B Table 7, findings are robust to this alternative modeling decision.

Finally, we examined sensitivity of findings to measurement decisions for state standardized test score outcomes (Appendix B Table 8). Our primary standardized test score outcomes are defined as the score the student received on the exam in a given grade the first time they took the test in that grade (column 1 in Appendix B Table 8, i.e., Score 1). To check that students retained in grade, or skipping grades, did not affect test score findings, we present three alternative score definitions in which we account for off-track students in three different ways. Score 2 removes any student taking a test in a different grade than expected – i.e., students who have been either retained or who skipped a grade. However, we allow students to rejoin the analytic sample in a given year, if they get back on track in that year. Score 3 removes a student from the analytic sample for a given grade and for all subsequent grades the first time they take a different grade-level test than the majority of their classmates. Score 4 removes all students from all years if they were ever retained in grade or skipped a grade. Findings are qualitatively similar across these approaches.

Note that in our findings through third grade (Weiland et al., 2020), we also examined robustness to several other threats to validity. In one check, we refit models with dichotomous outcomes using logistic regression instead of linear probability models. In another, we refit impact models with Pre-K enrollment defined as being enrolled in at least 150 days of the school year (instead of 1 day, as in our primary approach). Notably, there is no agreed-upon threshold in the literature and recent preschool studies have used different thresholds (Lipsey et al., 2018; Phillips et al., 2016). For parsimony and because prior findings were not sensitive to these alternative modeling choices, we did not repeat them in our present study.

External validity

As we have emphasized throughout, our lottery sample represents 24% of all applicants to Boston Pre-K in our focal years. Accordingly, the degree to which our findings apply to all students served is an important question. Weiland et al. (2020) previously showed that lottery sample children differ on multiple dimensions from the full sample of applicants. This includes background characteristics (see Table 1) – i.e., lottery sample children overall are proportionately more White and less likely to be eligible for free-reduced-lunch than the full sample. Weiland et al. (2020) also found that lotteries were highly concentrated in a subset of schools that had higher test scores and enrolled more White and non-free-reduced-lunch-eligible students than other district elementary schools. Further, 76% of the full sample who did not enroll in Boston Pre-K attended a non-BPS center-based preschool (vs. 97% of lottery control group members and 88% of lottery control group compliers). Preschool types also were markedly different than those in the lottery control group, with lottery control compliers more likely to be in private center and less likely to be in Head Start than full sample children who did not enroll in Boston Pre-K. Finally, Weiland et al. (2020), using a propensity-score approach, showed that although effects were null on grade retention, special education placement, and third grade test scores in the lottery sample, this did not appear to be case in the full sample. Rather, there were small but educationally meaningful associations between Pre-K enrollment and these outcomes in the full sample of Pre-K applicants. Specifically, they found that enrollees were 4 percentage points less likely to be retained in grades K-3 (p<0.001) and 7 percentage points less likely to have been placed in special education in grade K-3 (p<0.001). Compared to non-enrollees, enrollees also had third grade math and reading test scores that were 0.04 SDs higher than the average BPS third grader on both Math and English Language Arts (p<0.05).

In our extension of Weiland et al. (2020), we used the same propensity scores approach on our key outcomes in 4th-8th grades. To do so, we first predicted the probability that a student would be treated conditional on their background characteristics, their cohort year, and the public school each student lived closest to as a proxy for neighborhood characteristics. Then, we inverted these propensities to obtain an inverse probability weight (IPW) that we could use in our subsequent regression analysis to counteract selection into the program (Imbens & Wooldridge, 2009; Murnane & Willett, 2010). To gauge whether this approach captured selection into the program and following Abdulkadiroğlu and colleagues (2011), we began by replicating our lottery-based findings first. To do so, we fit IPW models within the sample of children who applied to Boston Pre-K via the four rounds of the lottery (N~ 9,971) and then within the full sample who applied, effectively adding in children who applied outside of the lottery process (N~12,740).

As we show in Table 6, our replication sample estimates suggest our IPW estimation approach reasonably recovers our lottery sample ITT impacts. Some replication sample estimates – like enrolled in BPS in 4-5 – are much larger in the replication sample than the lottery sample (14.0 pp vs. 8.5). Others, like 4th-8th standardized test scores, are smaller in magnitude in the replication sample. As in Weiland et al. (2020), full sample IPW estimates suggest larger benefits of Pre-K enrollment than lottery sample estimates. Interestingly, however, Algebra estimates are smaller in the full sample than the lottery or replication samples. We interpret these findings similarly to Weiland et al. (2020), as evidence that our lottery-based impact estimates may not apply to the full sample of Boston Pre-K applicants and enrollees.

Discussion

Rigorous, longitudinal evidence on modern-day Pre-K programs is a work in progress. In our extension of evidence on Boston Pre-K through the middle school years, we find substantial impacts on children's educational trajectories and some suggestive evidence of positive impacts on children's achievement, particularly on their access to and success in more advanced math courses. Some of our outcomes are novel to this literature and suggest new pathways to investigate in other contexts to advance our understanding of the conditions under which Pre-K effects may or may not be sustained. Our findings are particularly noteworthy given the relative high-quality of Boston Pre-K in these years, the unusually strong counterfactual in our study, the previous pattern of null results through third grade for treatment compliers, the developmental demands of the middle school years, and a similar pattern that suggests benefits in the full sample of applicants to the program, beyond the lottery-sample children.

Taking our key findings in turn, we found very large impacts of 29 percentage points (*p*<.01) on children's persistence in Boston Public Schools in late elementary school and of 18 percentage points (*p*<.05) in the middle school years for treatment compliers. Treatment compliers were also substantially less likely to enroll in charter schools in these years. Weiland et al. (2020) examined persistence in public schools as a Pre-K outcome in their study of Boston Pre-K through third grade because it was one of the City's original goals in offering the program. Beyond Boston, other localities and school systems have cited a similar rationale for offering Pre-K. For example, some New Orleans schools offer Pre-K to attract and retain families (Weixler et al., 2017) and DC chose school re-enrollment as a measure of school success/ progress under the Every Student Succeeds Act (U.S. Department of Education, 2017). On this metric, Pre-K programs in DC, Tulsa, and Boston appear to be succeeding (Gormley et al., 2018; Braga et al., 2024; Weiland et al., 2020). These findings are particularly notable as localities throughout the U.S. seek to address the current large post-COVID declines in public school enrollment (Goulas, 2024). Early entrance to public schools appears to generate a contextual school system "stickiness" that can last through middle school, though the benefits for students of doing so are unclear.

Notably, our findings also show novel differential navigation of accelerated curriculum opportunities within the public schools by treatment status as well, though differences were not always statistically significant. Specifically, we found that treatment compliers were less likely to be enrolled in third grade in schools that offered BPS's Advanced Work Course (AWC) program for qualifying fourth through sixth graders and subsequently, to enroll in AWC in 4^{th}-6^{th} grades. However, treatment compliers were more likely to take the test for Boston's sought-after 7^{th}-12^{th} grade competitive exam schools and were admitted and enrolled in exam schools at higher rates than children in the City overall. In 2007-2013, 35% of eligible students in Boston took the exam, 17% were admitted, and 15% enrolled (Goodman & Rucinski, 2018). A smaller share of control compliers in our study took the test (27%) but these students had similar success rates. In contrast, 42% of treatment compliers took the test, 26% were admitted, and 22% enrolled. The only other Pre-K follow-up study to our knowledge that has examined similar such outcomes in middle school is in Tulsa (Gormley et al., 2018). That study found that Tulsa Pre-K children but not Tulsa CAP Head Start children were more likely to be enrolled in honors courses in middle school; there was no difference for either in gifted program enrollment. Examining how Pre-K programs led to differential paths in crowded choice environments that offer a variety of accelerated curriculum opportunities may be key to understanding why Pre-K impacts appear to last in some contexts and not in others.

Regarding student progress and engagement, we continue to find no meaningful impacts of Boston Pre-K on in-grade retention or special education, consistent with Weiland et al. (2020), and with the addition of similar patterns for attendance. These findings are somewhat surprising, given the meta-analytic preschool evidence showing substantial benefits on these outcomes (McCoy et al., 2017) but they are consistent with some more recent Pre-K evaluations (Gormley et al., 2018; Gray-Lobe et al., 2024; Humphries et al., 2024; Phillips et al., 2016). Notably, although our lottery sample children were more advantaged with higher test scores and attendance than their BPS peers, their retention and special education rates were similar to district averages. Consistent with Weiland et al. (2020), we also found Pre-K enrollment was associated with benefits on these outcomes in the full sample of applicants. Understanding *how* schools make in-grade retention or special education decisions should be a priority for future research, given mixed findings for these outcomes in Pre-K studies and the relative high cost of these decisions for school systems.

For academic outcomes, our pattern of findings is notable. In 4^{th} grade, we found test score null impacts that were similar to Weiland et al.'s (2020) 3rd grade findings. However, in 5^{th} to 7th grade, the pattern of complier effects shows a pattern increasingly favoring treatment compliers, though only the impact on 7^{th} grade math (d=.24) is statistically significant. Interestingly, 7^{th} grade is the only school year in which we also found a statistically significant reduction in suspension rates of 7.6 percentage points. Seventh grade is also the year that exam schools in Boston start, in which a higher share of our treatment compliers enrolled. Further, it is the key year in which decisions are made that put children on a path (or not) to completing Algebra I by the end of 8^{th} grade (i.e., Pre-Algebra in 7^{th}, before possible Algebra I enrollment in 8^{th}). As described, we found very large, statistically significant effects on Algebra I enrollment and success by the end of 8^{th} grade for treatment compliers of 23 percent points.

In interpreting these findings, specifics of the Boston Pre-K program are important. Notably, the Boston Pre-K model stands out in its use of an evidence-based math curriculum that has shown large impacts on children's math skills at kindergarten entry in rigorous studies, including in Boston (Clements et al., 2011; 2013). In contrast, most other public Pre-K programs use "whole child" curricula and spend very little time on math specifically, especially beyond basic counting activities (Guerrero Rosada et al., 2025). Notably, a prior randomized trial of Building Blocks in the Pre-K to first grade years that was then followed by a quasi-experimental follow-up in fourth and fifth grades shows a pattern similar to the Boston context (Clements et al., 2023). In that study, there were large initial impacts of Building Blocks at the end of Pre-K, with lasting but smaller benefits through the end of first grade. There were no differences in fourth grade, but there was a rebound in treatment effects in fifth grade. The authors posited that their findings were due to latent foundation hypothesis – i.e., long term effects of early development are more detectable in periods in which curriculum complexity and expectations increase substantially. If so, early math skills may meet criteria for a fundamental, malleable skill (Bailey et al., 2017) for promoting impacts in later development periods.

However, we cannot disentangle achievement impacts from our educational trajectory findings. That is, schooling context may also have played a role. Notably, differences on test scores and suspension rates favoring treatment group children increase in the 6th grade, when they are substantially less likely to be in the youngest grade in their school (i.e., to have faced a transition to different school for 6th grade). As we reviewed, the literature has long highlighted that changing schools is out of sync with children's developmental needs during puberty (Eccles, 2004; Eccles & Wigfield, 1997). For math context specifically, as we described, we found that treatment compliers were in substantially more ambitious math instructional school environments in 8th grade, meaning higher proportions of 8th graders at their schools were enrolled in Algebra I (or higher) than for control compliers. The different educational trajectories for treatment compliers may have represented a "sustaining environment" (Bailey et al., 2017), echoing findings in the only randomized studies of this hypothesis, all of which found evidence for this hypothesis in explaining the persistence of Pre-K effects into the elementary school years (Clements et al., 2013; Mattera et al., 2021; Unterman & Weiland, 2024).

More broadly, taking Algebra I by the end of 8th grade has become of great policy and research interest, due to evidence from rigorous quasi-experimental research that taking more demanding math courses has many positive benefits for student achievement (Brummett et al., 2024) and postsecondary degree attainment (Nomi et al., 2021). Correlational research also shows that taking Algebra I by the end of 8th predicts higher earnings in adulthood (James, 2013; Joensen & Nielsen, 2009; Koedel & Tyhurst, 2012). Our findings are in need of replication in other contexts; to our knowledge, ours is the first longitudinal Pre-K study to examine this outcome. But our findings break new ground in pointing to possible benefits of Pre-K and to a possible new mechanism of Pre-K longitudinal effects (more advanced coursework), especially in light of the use of Boston's unusual use of an evidence-based math curriculum in Pre-K.

Our study has several important limitations. Measures of children's skills in the later grades are limited to reading and math state standardized tests and to only proxies for behavior. We also lacked AWC test results and admission offer information, preventing fuller understanding of that potential pathway. As we have acknowledged throughout, lottery sample children represent only 24% of all Boston Pre-K applicants in these years and differ from other applicants in some of their demographic characteristics. We conducted generalizability analysis following Weiland et al. (2020), but these are associations only. Ultimately, our findings have less external validity than we would like, both in Boston and outside it. We also have not yet followed these students into high school and beyond, though we are doing so in an in-progress study. In addition, new research has highlighted how preschool expansions can change the counterfactual (Berne et al., 2024), but we lack studies of these changes in Boston. Finally, as in Weiland et al. (2020), one assumption underlying our CACE analysis – that always-takers in both the treatment and control groups (i.e., children who would have enrolled in Boston Pre-K regardless of their treatment assignment status) experienced the same effect of enrollment – is difficult to evaluate.[11]

[11] Weiland et al. (2020) explained this issue in detail (p. 1414): "Treatment group always-takers in our study largely enrolled in their first-choice school; among control group crossovers, approximately a third did so. If the level at which a student ranked a prekindergarten program is indicative of their match with the program or its quality, it is

In close, the research is clear that Pre-K effects can last into the late elementary and middle school years and beyond but do not always, for reasons that are not well understood. Our study of Boston's high-quality Pre-K program into the often developmentally turbulent middle school years, compared against an unusually strong counterfactual, provides additional evidence on understanding these patterns.

possible that the two-thirds of the control group crossovers that enrolled in lower choices experienced a lower-quality program. Empirically, when we compared the first- choice schools and BPS schools actually attended in prekindergarten for the two-thirds of control group crossovers who did not enroll in their first-choice but enrolled in a lower choice, we found that their first-choice and their school attended differed on 9 out of 13 school context characteristics, with first-choice school appearing generally somewhat higher quality than the school in which they actually enrolled. However, we also found that school-level context variables were only weakly correlated with observed prekindergarten process quality (Weiland & Unterman, 2019). Further, among all control crossovers, about two-thirds were unassigned to the program after their first round; they were not assigned to a lower choice as part of the first round of the lottery system. These students crossed over later, which might indicate that their parents were particularly highly motivated and that therefore, they might have benefitted *more* from their Boston prekindergarten classroom. Ultimately, the direction of any potential bias from violation of the always-taker CACE assumption is ambiguous."

References

Abdulkadiroğlu, A., Angrist, A., Dynarski, S., Kane, T., & Pathak, P. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics*, *126*, 649-748.

Abdulkadiroglu, A., Angrist, J. D., Narita, Y., & Pathak, P. A. (2015). *Research design meets market design: Using centralized assignment for impact evaluation* (No. w21705). Cambridge, MA: National Bureau of Economic Research.

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, *82*, 137-196.

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, *34*, 275-318.

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

Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly*, *56*, 248-259.

Bernard, J., Lipsey, M., & Durkin, K. (2024) *Examining Pre-K effectiveness by alternative childcare options: Evidence from the Tennessee Voluntary Pre-K Study.* Washington, DC: Presentation at the Society for Research in Educational Effectiveness Conference, https://sree.confex.com/sree/2024/meetingapp.cgi/Paper/5744.

Berne, J. S. (2024). *The long-run impacts of Universal Pre-K with equilibrium considerations.* Providence, RI: Annenberg Working Paper.

Bloom, H. S., & Unterman, R. (2014). Can small high schools of choice improve educational prospects for disadvantaged students? *Journal of Policy Analysis and Management*, *33*, 290-319.

Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. SSRN 2594430.

Bloom, H. S., Zhu, P., & Unlu, F. (2010). *Finite sample bias from instrumental variables analysis in randomized trials*. MDRC Working Paper. New York: MDRC.

Boston Public Schools. (2013). *About Advanced Work Class.* https://www.bostonpublicschools.org/cms/lib07/ma01906464/centricity/domain/216/awc_sy14_pamphlet_eng.pdf

Bound, J., Jaeger, A., & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90*, 443-450.

Braga, B., Doromal, J. B., Greenberg, E., Monarrez, T., Restrepo, L., & Lamb, R. (2024). *The effects of public Pre-K for 3-year-olds on early elementary school outcomes: Evidence from the DC centralized lottery*. EdWorkingPaper No. 24-1019. Providence, RI: Annenberg Institute for School Reform at Brown University.

Brummet, Q., Liebert, L., Domina, T., Yoo, P. Y., & Penner, A. (2023). Early algebra affects peer composition. *Educational Evaluation and Policy Analysis*.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*, 684-698.

Cascio, E. U. (2021). Early childhood education in the United States: What, when, where, who, how, and why 1. In *The Routledge Handbook of the Economics of Education* (pp. 30-72). Routledge.

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

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2021). *Cradle to kindergarten: A new plan to combat inequality*. New York, NY: Russell Sage Foundation.

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

Clements, D. H., Sarama, J., Layzer, C., & Unlu, F. (2023). Implementation of a scale-up model in early childhood: Long-term impacts on mathematics achievement. *Journal for Research in Mathematics Education*, *54*, 64-88.

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

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal, 50*, 812 - 850. doi: 10.3102/0002831212469270.

Cicirelli, V. G. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development*. Athens, OH: Westinghouse Learning Corporation.

Cohodes, S. (2020). The long-run impacts of specialized programming for high-achieving students. *American Economic Journal: Economic Policy, 12*, 127-166.

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2016). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development*, doi:10.1111/cdev.12645

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development*, *88*, 996-1014.

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, *58*, 470.

Eccles, J. S. (2004). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology* (2nd ed., pp. 125–153). Hoboken, NJ: Wiley.

Eccles, J. S., & Wigfield, A. (1997). *Young adolescent development*. Columbus, OH: National Middle School Association.

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Iver, D. M., & Feldlaufer, H. (1993).

Negative effects of traditional middle schools on students' motivation. *The Elementary School Journal*, *93*, 553-574.

Feller, A., Grindal, T., Miratrix, L., & Page, L. (2016) Compared to what? Variation in the impact of early childhood education by alternative care type. *Annals of Applied Statistics*, *110*, 1245–1285.

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The BE Journal of Economic Analysis & Policy*, *8*.

Friedman-Krauss, A. H., Barnett, W. S., Hodges, K. S., Garver, K. A., Jost, T. M., Weisenfeld, G., Duer J. (2024). *The State of Preschool 2023: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Gennetian, L., Morris, P., Bos, J., & Bloom, H. (2005). Using instrumental variables analysis to learn more from social policy experiments. In H. Bloom (Ed.) *Learning More from Social Experiments: Evolving Analytic Approaches* (pp. 75-114). New York: Russell Sage.

Goodman, J., & Rucinski, M. (2018). *Increasing diversity in Boston's exam schools.* Cambridge, MA: Harvard Kennedy School Rappaport Institute for Greater Boston.

Gormley Jr, W. T., Phillips, D., & Anderson, S. (2018). The effects of Tulsa's pre-K program on middle school student performance. *Journal of Policy Analysis and Management*, *37*, 63-87.

Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, *31*, 392-415.

Goulas, S. (2024). *Breaking down enrollment declines in public schools*. Washington, DC: Brookings Institute.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*, 363-411.

Guerrero Rosada, P., Weiland, C., McComrick, M., Snow, C., Hsueh, J., Maier, M., & Sachs, J. (2025). *When time is relative: Pre-K classroom time use is measure-sensitive.* Working paper.

Heckman, J. (2000). Policies to foster human capital. *Research in Economics, 54,* 3–56 doi:10.1006/reec.1999.0225.

Hein, V., Smerdon, B., & Sambolt, M. (2013). *Predictors of postsecondary success*. College and Career Readiness and Success Center.

Hill, C. J., Gormley Jr, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, *32*, 60-79.

Humphries, J. E., Neilson, C., Ye, X., & Zimmerman, S. D. (2024). *Parents' earnings and the returns to universal pre-kindergarten* (No. w33038). Cambridge, MA: National Bureau of Economic Research.

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

James, J. (2013). The surprising impact of high school math on job market outcomes. *Economic Commentary*, 2013-14.

Joensen, J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, *44*, 171-198.

Kabay, S., Weiland, C., & Yoshikawa, H. (2020). Costs of the Boston public prekindergarten program. *Journal of Research on Educational Effectiveness*, *13*, 574-600.

Koedel, C., & Tyhurst, E. (2012). Math skills and labor-market outcomes: Evidence from a resume-based field experiment. *Economics of Education Review*, *31*, 131-140.

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, *45*, 155-176.

Ludwig, J., & Kling, J. (2007). Is crime contagious? *Journal of Law and Economics*, *50*, 491-518.

Marshall, N. L., Roberts, J., & Mills, L. (2006). *Boston Public Schools K1 and K2 programs needs assessment: Internal report to the Department of Early Childhood, Boston Public Schools.* Wellesley, MA: Wellesley Centers for Women.

Massachusetts Department of Elementary and Secondary Education. (2019). *2018-19 student attendance (Boston).* Malden, MA.

Massachusetts Department of Elementary and Secondary Education. (2024). *2024 retention report (Boston) – all students.* Malden, MA.

Massachusetts Department of Elementary and Secondary Education. (2025). *2024-2025 enrollment by selected population (Boston).* Malden, MA.

Massachusetts Department of Elementary and Secondary Education. (2016). *Spring 2016 district assessment decisions as of February 18, 2016.* Malden, MA. https://www.doe.mass.edu/news/news.aspx?id=21379

Massachusetts Department of Elementary and Secondary Education. (2022). *Researchers' guide to Massachusetts state education data.* Malden, MA.

Masten, A. S., & Cicchetti, D.(Eds.). (2010). Developmental cascades. *Development and Psychopathology*, *22,* 491-495.

Mattera, S. K., Jacob, R., MacDowell, C., & Morris, P. A. (2021). *Long-term effects of enhanced early childhood math instruction: The impacts of Making Pre-K Count and High 5s on third-grade outcomes*. New York, NY: MDRC.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., ... & Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher*, *46*, 474-487.

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

National Academies of Science, Education, and Medicine. (2024). *A new vision for high quality Pre-K curriculum.* Washington, DC.

National Center for Education Statistics. (2024). *Students with disabilities.* Washington, DC. https://nces.ed.gov/programs/coe/indicator/cgg/students-with-disabilities

National Mathematics Advisory Panel. (2008). *The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED500486.pdf

Nomi, T., Raudenbush, S.W., & Smith, J.J. (2021). Effects of double-dose algebra on college persistence and degree attainment, *Proceedings of the National Academy of Sciences, 118*, https://doi.org/10.1073/pnas.2019030118.

North, E. A., Ryan, A. M., Cortina, K., & Brass, N. R. (2019). Social status and classroom behavior in math and science during early adolescence. *Journal of Youth and Adolescence*, *48*, 597-608.

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

Phillips, D., Gormley, W., & Anderson, S. (2016). The effects of Tulsa's CAP Head Start program on middle-school academic outcomes and progress. *Developmental Psychology*, *52*, 1247.

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

Rochester, S. E., Weiland, C., Unterman, R., McCormick, M., & Moffett, L. (2019). The little kids down the hall: Associations between school climate, pre-K classroom quality, and pre-K children's gains in receptive vocabulary and executive function. *Early Childhood Research Quarterly*, *48*, 84-97.

Ryan, A. M., Shim, S. S., & Makara, K. A. (2013). Changes in academic adjustment and relational self-worth across the transition to middle school. *Journal of Youth and Adolescence*, *42*, 1372-1384.

Sameroff, A. (2009). *The transactional model of development: How children and contexts shape each other.* Washington, D.C.: American Psychological Association.

Sawhill, I. V., & Welch, M. (2023*). Early childhood education, What do we know and what should we do?* Washington, DC: Brookings Institute.

Schickedanz, J., & D. Dickinson. (2005). *Opening the World of Learning*. Iowa City, IA: Pearson Publishing.

Schochet, P. Z. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review*, *33*(6), 539-567.

Schweinhart, L. J., & Weikart, D. P. (1981). Effects of the Perry Preschool Program on youths through age 15. *Journal of the Division for Early Childhood*, *4*, 29-39.

Shapiro, A., Martin, E., Weiland, C., & Unterman, R. (2019). If you offer it, will they come? Patterns of application and enrollment behavior in a universal prekindergarten context. *AERA Open*, *5*(2), 2332858419848442.

Unterman, R., & Weiland, C. (2024). Quantifying and predicting variation in the medium-term effects of an oversubscribed prekindergarten program. *Journal of Research on Educational Effectiveness*, 1-28.

U.S. Department of Education. (2017). *District of Columbia revised state template for the consolidated state plan: The Elementary and Secondary Education Act of 1965, as amended by the Every Student Succeeds Act*. https://osse.dc.gov/sites/default/files/dc/sites/osse/documents/OSSE%20ESSA%20State%20Plan_%20May%202%202017.pdf.

Warren, J. R., Hoffman, E., & Andrew, M. (2014). Patterns and trends in grade retention rates in the United States, 1995–2010. *Educational Researcher*, *43*, 433-443.

Weiland, C. (2016). Launching preschool 2.0: A road map to high-quality public programs at scale. *Behavioral Science & Policy*, *2*, 37-46.

Weiland, C., & Rosada, P. G. (2022). *Widely used measures of Pre-K classroom quality: What we know, gaps in the field, and promising new directions*. New York, NY: MDRC.

Weiland, C., Sachs, J., McCormick, M., Hsueh, J., & Snow, C. (2021). Fast-response research to answer practice and policy questions. *The Future of Children*, *31*, 75-96.

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

Weiland, C., Unterman, R., & Shapiro, A. (2021). The kindergarten hotspot: Literacy skill convergence between Boston prekindergarten enrollees and nonenrollees. *Child Development*, *92*, 600-608.

Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. *Child Development*, *91*, 1401-1422.

Weiland, C. & Yoshikawa, H. (2013). The impacts of an urban public prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills: Evidence from Boston. *Child Development, 84*, 2112-2130.

Weiland, C., & Yoshikawa, H. (2014). Does higher peer socio-economic status predict children's language and executive function skills gains in prekindergarten? *Journal of Applied Developmental Psychology*, *35*, 422-432.

Weixler, L. B., Lincove, J. A., & Gerry, A. (2017). *The provision of public pre-k in the absence of centralized school management*. New Orleans, LA: Education Research Alliance for New Orleans Technical Report.

What Works Clearinghouse. (2014). *WWC procedures and standards handbook (Version 3.0).* Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse.

Whitehurst, G., & Klein, E. (2015). *Do we already have universal preschool?* Washington, DC: Economic Studies at Brookings.

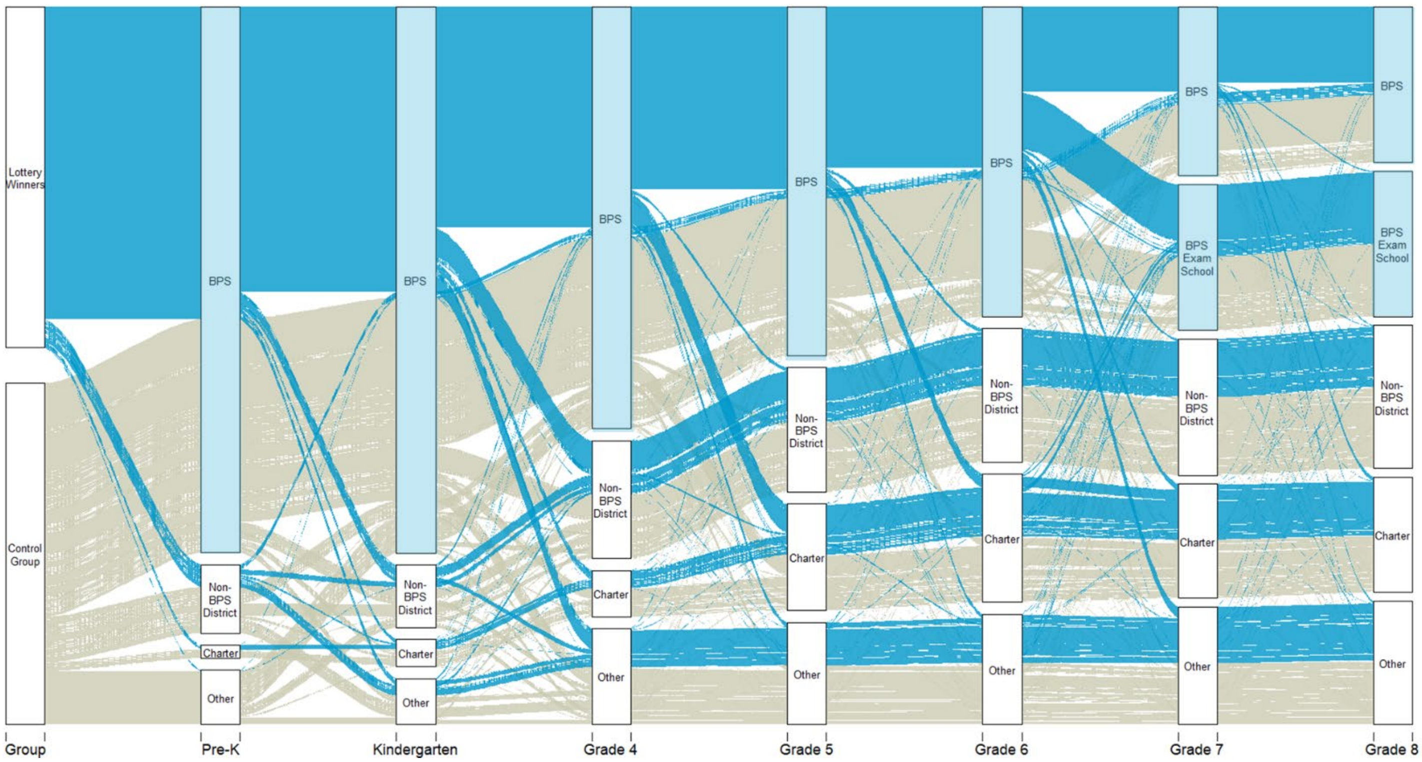
Wu, T., Weiland, C., Diemer, M., Unterman, R., & Shapiro, S. (2025). *The psychometrics of soft skills over time: Building longitudinal student behavior composites using administrative data*. Manuscript in preparation.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, 21-35.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* New York, NY: Foundation for Child Development, Society for Research in Child Development.

Figures and Tables

FIGURE 1: School Enrollment Pathways for Lottery Winners and Control Group Students through 8th grade



Note: This figure represents school enrollment pathways across different school types (BPS, BPS Exam School, non-BPS district school, charter school, and other) for both lottery winners and the control group. The thickness of the lines represents the percentage of lottery winners/control group students that are enrolled in each type of school. Blue lines show the pathways taken by lottery winners, and BPS-specific school categories are highlighted in the light blue boxes. BPS = Boston Public Schools.

TABLE 1: Balance on observables in the first-choice lottery sample

	Lottery winners	Control group	Estimated difference	*P*-value	All Pre-K Applicants
Race/ethnicity (%)					
Hispanic	35.54	41.26	-5.72***	0.002	43.87
Black	23.82	22.29	1.53	0.309	28.43
White	27.60	25.05	2.55	0.120	17.06
Asian	10.11	7.26	2.85***	0.004	7.58
Other	2.93	4.15	-1.22	0.146	3.06
Male (%)	50.09	46.86	3.23	0.137	51.72
Eligible for free/reduced lunch (%)	56.71	55.85	0.86	0.650	65.07
Age	4.51	4.53	-0.02	0.119	4.52
Country of origin USA (%)	95.08	94.72	0.36	0.699	93.33
Home language (%)					
English	51.98	54.96	-2.98	0.120	56.68
Spanish	25.42	25.47	-0.05	0.977	24.36
Other	22.60	19.57	3.03~	0.061	18.95
N children	1,063	2,029			12,740

Note: For the lottery sample, there was a small amount of missing data on all baseline characteristics except age: 11 children (0.36%) were missing race/ethnicity and male information, 32 (1.03%) were missing male and free/reduced lunch information, 110 (3.56%) were missing country of origin information, and 4 (0.13%) were missing home language information. Means in the table were computed using non-missing data. Values for first choice lottery winners are the simple means for each variable. Values for the difference between lottery winners and control group members are obtained from a regression of a given baseline characteristic on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on lottery indicator equals the difference in the mean baseline characteristic between lottery winners and control group members, respectively. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated differences. An F-test was used to assess the statistical significance of the overall difference between lottery winners and control group members reflected by the full set of baseline characteristics in the table. The resulting F value is not statistically significant (p=0.213). Statistical significance levels are indicated as: ***$p < .001$; ~$p < .10$.

TABLE 2: Educational trajectory impacts

	ITT	Standard Error	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
Enrolled in BPS (%)						
4th grade	8.97***	2.88	30.80**	8.76	72.03	41.23
5th grade	7.75**	2.65	26.59**	8.33	60.06	33.47
6th grade	7.14**	2.88	24.52**	9.20	53.66	29.14
7th grade	5.91*	2.69	20.28*	8.66	52.12	31.84
8th grade	5.54*	2.63	19.03*	8.46	49.54	30.51
Enrolled 4-5	8.47**	2.60	29.08**	8.09	59.55	30.47
Enrolled 6-8	5.14~	2.65	17.63*	8.72	44.17	26.54
Enrolled in a charter school (%)						
4th grade	-4.17**	1.38	-14.06**	4.33	4.51	18.57
5th grade	-2.74~	1.73	-9.20~	5.43	16.02	25.22
6th grade	-2.11	1.85	-7.19	5.87	19.88	27.07
7th grade	-2.39	1.75	-8.28	5.40	18.82	27.10
8th grade	-1.68	1.61	-5.87	5.10	20.39	26.26
Enrolled 4-5	-3.63**	1.22	-12.25**	3.76	4.23	16.48
Enrolled 6-8	-1.89	1.80	-6.45	5.70	24.43	30.88
Accelerated coursework (%)						
Enrolled in AWC-offering BPS school (3rd grade)	-3.59	2.26	-12.31~	7.38	19.50	31.81
Ever enrolled in AWC 4th-6th	-3.48*	1.45	-11.93*	4.85	11.25	23.18
Exam schools (%)						
Took exam school test in 6th-8th	4.18~	2.35	14.36~	7.47	41.69	27.33
Admitted to exam school 7th-8th	3.04	2.52	10.44	7.96	25.59	15.15
Enrolled in exam school 7th-8th	1.44	2.22	4.95	7.14	21.88	16.93

Note: There was no missing data on enrollment and exam school variables. Other outcomes were missing data as follows: charter school enrollment: 11-23% across variables; and Advanced Work Course: 1-5% across variables. Statistical significance levels are indicated as: ~p<0.10; *p < .05; **p < .01; ***p < .001. Statistical significance indicators for ITT and CACE estimates on some variables do not perfectly match due to small rounding differences across estimation strategies.

TABLE 3: CACE Estimates of K-5 and 6-8 School Context Differences Between Lottery Winners and Control Group

	K-5 School Context				6-8 School Context			
	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
Student background characteristics								
% Low-income (K-5)/High needs (6-8)	-3.22	3.25	65.14	68.39	4.76	2.77	58.38	53.62
% English learners	3.58	3.47	27.53	23.95	4.62*	1.93	17.34	12.72
% non-English home language	5.35	4.37	39.22	33.87	5.69*	2.55	37.21	31.52
% Disabilities	0.43	1.24	17.67	17.24	0.78	1.28	17.28	16.50
% African American	-8.50*	3.94	25.87	34.37	-2.83	2.19	26.09	28.92
% Asian	-2.78~	1.34	8.37	11.15	-0.86	1.32	10.92	11.78
% Hispanic	10.41~	5.70	39.10	28.69	10.57**	2.56	31.77	21.20
% White	1.94	4.11	23.43	21.49	-7.00*	3.44	27.56	34.56
% Female	-0.34	0.68	48.27	48.61	0.46	0.49	49.80	49.34
Student performance								
ELA % Advanced or Proficient	-1.50	2.53	47.98	49.48	-4.82	3.01	46.12	50.94
% Advanced	-2.38	1.05	9.06	11.44	-1.40	1.19	8.51	9.91
% Proficient	1.12	1.85	39.15	38.03	-3.42	2.04	37.60	41.02
Math % Advanced or Proficient	0.76	2.47	50.68	49.92	-5.95~	3.03	43.32	49.27
% Advanced	-1.64	1.56	16.02	17.66	-2.77*	1.14	7.37	10.14
% Proficient	3.08~	1.71	34.95	31.87	-3.17	2.33	35.96	39.13
Teacher and school characteristics								
% Licensed teachers	3.93*	1.23	96.19	92.26	2.35	1.68	91.49	89.14
Student-teacher ratio	0.37	0.43	13.89	13.52	1.02~	0.51	14.01	12.99
% Exemplary teachers	4.41*	2.00	17.17	12.76	6.63***	2.02	20.79	14.16
% Proficient teachers	-2.92	2.16	77.19	80.11	-6.38***	2.13	74.49	80.87
% Teacher retention	0.76	1.53	79.58	78.82	0.24	1.53	78.91	78.67
Stability	2.01*	0.98	89.77	87.76	-0.35	0.79	91.39	91.74
Class Size	0.12	0.61	19.31	19.19	0.23	0.66	20.56	20.33

Note. Using publicly available data from the Massachusetts Department of Elementary and Secondary Education, we averaged available school-level data across the schools in which a student was enrolled for the longest period of time each year in K-5. If data were missing for a student in a given year (e.g., first grade), we used non-missing data to compute the student's K-5 context averages (e.g., K, second, third, fourth, fifth). Across K-5 variables, data were missing for 4%–10% of treatment students and 9-19% of control students on student and teacher characteristics. Across grade 6-8 variables, data were missing for 12%–22% of treatment students and 17-25% of control students on student and teacher characteristics. Percentage of teachers scoring proficient or exemplary on state ratings and average class size was available from 2014-2015 onward. Percentage of teachers retained, percentage of teachers licensed, and student-teacher ratio were missing from ~20% of schools in 2015-2016 and 2016-17, MA data collection changed measures for student poverty during the study period; K-5 poverty measure was % low-income and grade 6-8 poverty measure was % high needs for study cohorts 1-4. Statistical significance levels are indicated as: ~ p<0.10; *p < .05; **p < 01;***p < .001. ELA = English Language Arts.

TABLE 4: School progress/engagement impacts

	ITT	Standard Error	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
Retained in grade (%)						
3rd grade	-0.05	0.52	-0.17	1.69	1.21	1.38
4th grade	-0.15	0.28	-0.51	0.92	0.68	1.19
5th grade	-0.18	0.22	-0.62	0.72	0.23	0.85
6th grade	0.31	0.38	1.07	1.26	0.95	---
7th grade	-0.04	0.32	-0.14	1.06	0.97	1.11
Ever retained 3-7	-0.10	0.64	-0.35	2.09	3.85	4.20
Special Education Classification (%)						
4th grade	0.25	1.76	0.83	5.59	18.08	17.25
5th grade	0.72	1.94	2.42	6.23	18.85	16.43
6th grade	1.25	1.82	4.26	5.95	19.71	15.45
7th grade	1.22	1.96	4.17	6.48	20.07	15.90
8th grade	1.22	1.75	4.27	5.86	21.07	16.80
Ever SPED 4-8	1.33	1.77	4.48	5.79	21.71	17.23
Attendance Rate (%)						
4th grade	-0.09	0.23	-0.30	0.75	95.80	96.10
5th grade	-0.45*	0.20	-1.52*	0.67	95.89	97.41
6th grade	-0.45~	0.24	-1.52*	0.76	95.54	97.06
7th grade	-0.48*	0.19	-1.66**	0.62	95.18	96.84
8th grade	-0.07	0.30	-0.25	0.98	94.77	95.02
Average 4-8	-0.12	0.21	-0.41	0.68	95.40	95.52
Ever Suspended						
4th grade	-0.50	0.62	-1.69	2.01	2.23	3.92
5th grade	0.34	0.73	1.15	2.33	4.29	3.14
6th grade	-0.80	0.90	-2.74	2.89	6.30	9.04
7th grade	-2.19*	1.03	-7.57*	3.50	6.52	14.09
8th grade	0.21	1.36	0.73	4.47	7.24	6.51
Ever suspended 4-8	-1.60	1.32	-5.46	4.22	15.31	20.77

Note: Outcomes were missing data as follows: grade retention 7-19% across variables; special education 3-21% across variables; attendance rate 9-23% across variables; suspension 9-22% across variables. "--" indicates the control group complier mean is inestimable due to an imprecise CACE. Statistical significance levels are indicated as: ~ p<0.10; * p < .05; ** p < .01;*** p < .001.

TABLE 5: Academic achievement impacts

	ITT	Standard Error	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
State reading test scores						
4th grade	0.02	0.04	0.05	0.13	0.41	0.36
5th grade	0.03	0.04	0.10	0.13	0.39	0.29
6th grade	0.05	0.05	0.17	0.15	0.46	0.29
7th grade	0.05	0.04	0.17	0.13	0.40	0.23
8th grade^{+}	0.04	0.05	0.14	0.15	0.42	0.29
State math test scores						
4th grade	-0.00	0.06	-0.00	0.17	0.37	0.38
5th grade	0.03	0.04	0.09	0.12	0.38	0.30
6th grade	0.05	0.04	0.18	0.14	0.46	0.28
7th grade	0.07*	0.03	0.24*	0.12	0.43	0.20
8th grade^{+}	0.01	0.05	0.04	0.14	0.45	0.41
Algebra I (%)						
Took Algebra I by 8th grade	6.55**	2.44	22.64**	7.98	68.77	46.13
Passed Algebra I by 8th grade	6.52*	2.55	22.57**	8.06	61.81	39.24

Note: Outcomes were missing data as follows: state reading test scores 15-25% across variables from 4th to 7th grade and 42-46% for 8th grade; state math test scores 15-25% across variables from 4th to 7th grade and 43-46% for 8th grade; and Algebra 17-24% across variables. Statistical significance levels are indicated as: ***=.1 percent ** = 1 percent; * = 5 percent.

$^{+}$Data are missing for most of cohort 4 in 8th grade due to the cancellation of state testing in spring 2020.

TABLE 6. Lottery ITT results and IPW results for the lottery replication sample and the full Boston Pre-K applicant sample

	Lottery Sample		Replication Sample		Full Sample	
	ITT	*SE*	ITT	*SE*	ITT	*SE*
Enrolled in BPS (%)						
BPS Enrolled 4-5	8.47**	2.60	14.04***	1.18	8.08***	0.97
BPS Enrolled 6-8	5.14~	2.65	8.97***	1.14	5.56***	0.94
Enrolled in a Charter School (%)						
Charter Enrolled 4-5	-3.63**	1.22	-6.14***	0.80	-2.63***	0.59
Charter Enrolled 6-8	-2.71	1.84	0.61	1.13	1.72	0.92
Exam Schools (%)						
Took exam school test in 6th-8th	4.19~	2.35	3.22**	1.11	5.20***	0.87
Admitted to exam school 7th-8th	3.04	2.52	3.56***	0.87	4.46***	0.68
Enrolled in exam school 7th-8th	1.44	2.22	3.83***	0.82	4.62***	0.64
School Progress/Engagement						
Ever SPED 4-8	1.33	1.77	-1.16	1.01	-4.17***	0.85
Ever retained 3-7	-0.10	0.64	0.91	0.48	-0.00	0.43
Average Attendance Rate (4-8)	-0.12	0.21	-0.09	0.12	0.24*	0.10
Ever suspended (4-8)	-1.60	1.32	-1.52	1.02	-1.72*	0.83
State Reading Test Scores						
4th grade	0.02	0.04	-0.02	0.02	0.06**	0.02
5th grade	0.03	0.04	-0.01	0.02	0.06**	0.02
6th grade	0.05	0.05	0.01	0.02	0.07***	0.02
7th grade	0.05	0.04	-0.01	0.02	0.04*	0.02
8th grade	0.04	0.05	-0.03	0.03	0.02	0.02
State Math Test Scores						
4th grade	0.00	0.06	0.01	0.02	0.08***	0.02
5th grade	0.03	0.04	-0.01	0.02	0.06***	0.02
6th grade	0.05	0.04	0.01	0.02	0.07***	0.02
7th grade	0.07*	0.03	-0.01	0.02	0.05*	0.02
8th grade	0.01	0.05	-0.01	0.03	0.04	0.02
Algebra I Outcomes						
Took Algebra I by 8th grade	6.55**	2.44	5.18***	1.31	2.61*	1.05
Passed Algebra I by 8th grade	6.52*	2.55	4.22**	1.34	2.97**	1.09

Note: For the lottery sample (ITT), outcomes were missing data as follows: 0% for BPS enrollment outcomes; 11-19% for charter school outcomes; 0% for exam school outcomes; 3-14% for school progress/engagement; 17-24% for Algebra outcomes; 15-25% for test scores in 4^{th}-7^{th} grades; and 42-46% for 8^{th} grade test scores (due to COVID-19 closures when the final cohort was mostly in 8^{th} grade). For the replication sample, outcomes were missing data as follows: 0% for BPS enrollment outcomes; 8-16% for charter school outcomes; 0% for exam school outcomes; 1-13% for school progress/engagement; 13-18% for Algebra outcomes across variables; 12-36% for test score outcomes across variables. For the full sample, outcomes were missing data as follows: 0% for BPS enrollment outcomes; 10-25% for charter school outcomes; 0% for exam school outcomes; 1-21% for school progress/engagement; 17-30% for Algebra outcomes; 14-51% for test score outcomes. Statistical significance levels are indicated as: ~p<0.10; * p < .05; ** p < .01; *** p < .001. ITT = intent-to-treat; IPW = inverse probability weight; BPS = Boston Public Schools; SPED = special education.

APPENDIX A: Attrition analysis

Table 1: Percent of missing data for educational trajectory outcomes

	Lottery Winners	Control Group	Estimated Difference	*P*-value
Enrolled in a charter school (%)				
4th grade	11.76	20.80	-9.04	0.14
5th grade	13.17	21.14	-7.97	0.41
6th grade	14.77	21.93	-7.16	0.83
7th grade	17.03	21.83	-4.81	0.35
8th grade	17.87	22.67	-4.80	0.49
Enrolled 4-5	11.01	19.17	-8.17	0.29
Enrolled 6-8	12.70	17.79	-5.09	0.56
School Context (K-5)				
% Teachers rated exemplary	10.07	19.02	-8.96*	0.03
Student-teacher ratio	3.76	8.87	-5.11	0.84
Average class size	4.99	10.35	-5.36	0.93
% Low Income	3.95	9.31	-5.36	0.59
% Hispanic	3.76	8.77	-5.01	0.93
% Advanced or Proficient on ELA state	8.56	15.82	-7.26	0.25
% Advanced or Proficient on Math state	8.56	15.82	-7.26	0.25
School Context (6-8)				
% Teachers rated exemplary	21.54	25.58	-4.04	0.532
Student-teacher ratio	13.73	18.43	-4.70	0.242
Average class size	12.70	17.84	-5.14	0.610
% High Needs	12.70	17.79	-5.09	0.555
% Hispanic	12.70	17.79	-5.09	0.555
% Advanced or Proficient on ELA state	12.98	18.14	-5.15	0.542
% Advanced or Proficient on Math state	12.98	18.14	-5.15	0.542

Note: BPS enrollment and persistence, accelerated coursework (AWC), and exam school variables are not shown in the table because there were no missing data for these outcomes. Binary missing data indicators were generated for each variable, set to 1 if the student was not observed in the data or the student was observed but the value for the given variable was missing from their student records. Values for the percent missing for lottery winners are the simple means for each requisite variable. Values for the difference between lottery winners and control group members are obtained from a regression of the given missing variable indicator on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the percent missing for each variable between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated differences. Statistical significance levels are indicated as: *$p < .05$

Table 2: Percent of missing data for school progress/engagement outcomes

	Lottery Winners	Control Group	Estimated Difference	*P*-value
Retained in grade (%)				
3rd grade	7.62	12.86	-5.24	0.79
4th grade	8.84	13.80	-4.96	0.42
5th grade	10.16	14.74	-4.58	0.26
6th grade	11.76	15.87	-4.11	0.19
7th grade	14.11	18.97	-4.86	0.26
Ever retained 3-7	7.15	12.03	-4.88	0.72
Special Education Classification (%)				
4th grade	9.60	18.04	-8.44~	0.08
5th grade	11.19	18.88	-7.68	0.40
6th grade	12.04	19.32	-7.28	0.46
7th grade	13.45	18.04	-4.59	0.50
8th grade	16.18	20.80	-4.62	0.44
Ever SPED 4-8	2.63	6.51	-3.87	0.46
Attendance Rate (%)				
4th grade	11.76	20.80	-9.04	0.14
5th grade	13.17	21.14	-7.97	0.41
6th grade	14.77	21.93	-7.16	0.83
7th grade	17.12	21.93	-4.81	0.35
8th grade	17.87	22.67	-4.80	0.49
Average 4-8	9.41	14.49	-5.08	0.56
Ever Suspended				
4th grade	11.10	19.62	-8.51	0.24
5th grade	12.32	20.11	-7.78	0.42
6th grade	14.02	21.04	-7.03	0.84
7th grade	15.80	20.80	-4.99	0.51
8th grade	16.56	21.54	-4.98	0.62
Ever suspended K-8	9.31	14.49	-5.18	0.61

Note: Binary missing data indicators were generated for each variable, set to 1 if the student was not observed in the data or the student was observed but the value for the given variable was missing from their student records. Values for the percent missing for lottery winners are the simple means for each requisite variable. Values for the difference between lottery winners and control group members are obtained from a regression of the given missing variable indicator on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the percent missing for each variable between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated differences. Statistical significance levels are indicated as: ~p<0.10

Table 3: Percent of missing data for academic achievement outcomes

	Lottery Winners	Control Group	Estimated Difference	*P*-value
State reading test scores				
4th grade	15.05	22.97	-7.92	0.48
5th grade	15.71	22.77	-7.06	0.88
6th grade	17.59	24.35	-6.76	0.79
7th grade	20.13	24.35	-4.22	0.50
8th grade+	42.43	45.54	-3.11	0.50
State math test scores				
4th grade	14.77	22.72	-7.95	0.50
5th grade	15.90	23.22	-7.32	0.70
6th grade	17.87	24.24	-6.37	0.81
7th grade	20.51	24.50	-3.99	0.27
8th grade+	42.52	45.69	-3.17	0.56
Algebra I outcomes				
Took Algebra I by 8th grade	17.22	22.57	-5.36	0.77
Passed Alegra I by 8th grade	18.81	23.90	-5.09	0.67

Note: +Most of the last cohort in our study was in 8th grade when the COVID-19 pandemic began in spring 2020. Data are missing at much higher rates for 8th grade because state standardized tests were cancelled in spring 2020. Binary missing data indicators were generated for each variable, set to 1 if the student was not observed in the data or the student was observed but the value for the given variable was missing from their student records. Values for the percent missing for lottery winners are the simple means for each requisite variable. Values for the difference between lottery winners and control group members are obtained from a regression of the given missing variable indicator on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the percent missing for each variable between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated differences.

Table 4: Balance on baseline characteristics for students with 7th grade standardized test score data

	Lottery winners	Control group	Estimated difference	*P*-value
Race/ethnicity (%)				
Hispanic	35.49	40.15	-4.66*	0.03
Black	24.56	22.14	2.42	0.16
White	26.20	25.43	0.77	0.67
Asian	10.93	8.62	2.30*	0.05
Other	2.82	3.66	-0.84	0.37
Male (%)	48.53	46.27	2.26	0.36
Eligible for free/reduced lunch (%)	59.69	59.69	0.00	--
Age	4.51	4.54	-0.03	0.07~
Country of origin USA (%)	95.65	94.97	0.68	0.49
Home language (%)				
English	52.53	54.41	-1.89	0.39
Spanish	25.03	25.22	-0.19	0.91
Other	22.44	20.37	2.08	0.26
N children	851	1,539		

Note: 14 students were missing free/reduced price lunch information and 4 students were missing information for country of origin; all other data was available for all students. Values for lottery winners are the simple means for each requisite group. Values for the difference between lottery winners and control group members are obtained from a regression of a given baseline characteristic on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the mean baseline characteristic between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated difference. Statistical significance levels are indicated as: ~p<.10; *p<.05. An F-test was used to assess the statistical significance of the overall difference between lottery winners and control group members reflected by the full set of baseline characteristics in the table. The resulting F value was not statistically significant (p=0.220).

Table 5: Balance on baseline characteristics for students with any 4th-8th grade standardized test score data

	Lottery winners	Control group	Estimated difference	P-value
Race/ethnicity (%)				
Hispanic	36.11	40.56	-4.46*	0.02
Black	24.00	22.67	1.33	0.41
White	26.74	25.39	1.34	0.44
Asian	10.21	7.66	2.55*	0.02
Other	2.95	3.72	-0.77	0.38
Male (%)	50.00	47.36	2.64	0.26
Eligible for free/reduced lunch (%)	58.95	58.76	0.18	0.93
Age	4.51	4.54	-0.02~	0.09
Country of origin USA (%)	95.58	94.85	0.73	0.44
Home language (%)				
English	52.95	55.04	-2.1	0.31
Spanish	25.68	25.56	0.13	0.94
Other	21.37	19.4	1.97	0.25
N children	950	1,717		

Note: 15 students were missing free/reduced price lunch information and 4 students were missing information for country of origin; all other data was available for all students. Values for lottery winners are the simple means for each requisite group. Values for the difference between lottery winners and control group members are obtained from a regression of a given baseline characteristic on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the mean baseline characteristic between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated difference. Statistical significance levels are indicated as: ~$p<0.1$, *$p<.05$. An F-test was used to assess the statistical significance of the overall difference between lottery winners and control group members reflected by the full set of baseline characteristics in the table. The resulting F-value was not statistically significant ($p=0.224$).

Table 6: Balance on baseline characteristics for students with non-missing data for taking Algebra I by the end of 8th grade

	Lottery winners	Control group	Estimated difference	*P*-value
Race/ethnicity (%)				
Hispanic	36.70	40.61	-3.90~	0.06
Black	24.32	22.74	1.58	0.35
White	25.45	24.67	0.78	0.66
Asian	10.80	8.13	2.66*	0.02
Other	2.73	3.85	-1.12	0.23
Male (%)	48.52	46.53	1.99	0.41
Eligible for	59.89	60.16	-0.27	0.90
Age	4.52	4.54	-0.02	0.12
Country of	95.45	94.93	0.52	0.60
Home language (%)				
English	51.36	53.88	-2.52	0.24
Spanish	26.25	26.13	0.12	0.95
Other	22.39	19.98	2.40	0.19
N children	880	1,571		

Note: 15 students were missing free/reduced price lunch information and 4 students were missing information for country of origin; all other data was available for all students. Values for lottery winners are the simple means for each requisite group. Values for the difference between lottery winners and control group members are obtained from a regression of a given baseline characteristic on a series of indicator variables that identify each lottery plus a treatment indicator set to 1 for lottery winners and 0 for lottery losers. The coefficient on the treatment indicator equals the difference in the mean baseline characteristic between lottery winners and control group members. The value for control group members equals the corresponding value for lottery winners minus the estimated difference between lottery winners and control group members. A two-tailed t-test was applied to the estimated difference. Statistical significance levels are indicated as: ~$p<0.1$, *$p<.05$. An F-test was used to assess the statistical significance of the overall difference between lottery winners and control group members reflected by the full set of baseline characteristics in the table. The resulting *F*-value was not statistically significant ($p=0.201$).

APPENDIX B: Additional tables and robustness checks

Table 1: Lottery sample by year

Cohort year	N Applicants	N Lottery Participants	% of Applicants in Lotteries	N Schools	N Schools with Lotteries	% of Schools with Lotteries	N (%) schools 75% of lottery students competed for
Spring 2007	2,182	622	29	60	35	58	10 (17%)
Spring 2008	2,333	754	32	64	42	66	12 (19%)
Spring 2009	2,613	826	32	63	44	70	16 (25%)
Spring 2010	2,843	890	31	64	49	77	19 (30%)

Note: The total number of applicants to the four rounds of the lottery across the four years was 9,971. The total number of lottery participants was 3,092, 24% of the total number of applicants.

Table 2: ITT Estimates of K-5 and 6-8 School Context Differences Between Lottery Winners and Control Group

	K-5 School Context		6-8 School Context	
	ITT	Standard Error	ITT	Standard Error
Student background characteristics				
% Low-income (K-5)/High needs (6-8)	-0.97	1.01	1.39	0.85
% English learners	1.08	1.12	1.35*	0.60
% non-English home language	1.61	1.40	1.66*	0.80
% Disabilities	0.13	0.39	0.23	0.39
% African American	-2.57*	1.24	-0.83	0.68
% Asian	-0.84~	0.44	-0.25	0.41
% Hispanic	3.14~	1.83	3.09***	0.77
% White	0.58	1.30	-2.04*	1.04
% Female	-0.10	0.21	0.13	0.15
Student performance				
ELA % Advanced or Proficient	-0.45	0.82	-1.42	0.96
% Advanced	-0.71	0.35	-0.41	0.38
% Proficient	0.33	0.57	-1.01	0.65
Math % Advanced or Proficient	0.23	0.77	-1.75~	0.97
% Advanced	-0.49	0.51	-0.82*	0.38
% Proficient	0.92~	0.51	-0.94	0.73
Teacher and school characteristics				
% Licensed teachers	1.19*	0.39	0.69	0.53
Student-teacher ratio	0.11	0.13	0.30~	0.16
% Exemplary teachers	1.28*	0.59	1.98***	0.60
% Proficient teachers	-0.85	0.64	-1.91***	0.62
% Teacher retention	0.23	0.49	0.07	0.47
Stability	0.61*	0.30	-0.10	0.25
Class Size	0.04	0.19	0.07	0.20

Note: Using publicly available data from the Massachusetts Department of Elementary and Secondary Education, we averaged available school-level data across the schools in which a student was enrolled for the longest period of time each year in kindergarten, first, second, third, fourth, and fifth grade. If data were missing for a student in a given year (e.g., first grade), we used non-missing data to compute the student's K-5 context averages (e.g., K, second, third, fourth, fifth). Across K-5 variables, data were missing for 4%–10% of treatment students and 9-19% of control students on student and teacher characteristics. Across grade 6-8 variables, data were missing for 12%–22% of treatment students and 17-25% of control students on student and teacher characteristics. Percentage of teachers scoring proficient or exemplary on state ratings and average class size was available from 2014-2015 onward. Percentage of teachers retained, percentage of teachers licensed, and student-teacher ratio were missing from ~20% of schools in 2015-2016 and 2016-17. Massachusetts data collection changed measures for student poverty during the study period; K-5 poverty measure was % low-income and grade 6-8 poverty measure was % high needs for study cohorts 1-4. Statistical significance levels are indicated as: ~p<0.10; *p < .05; **p < .01;***p < .001. ELA = English Language Arts.

Table 3: ITT and CACE Estimates of 5th Grade School Context Differences Between Lottery Winners and Control Group

	ITT	Standard Error	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
Student background characteristics						
% Low-income (K-5)	5.21*	2.60	15.30*	8.18	65.17	49.87
% English learners	1.16	0.84	3.91	2.58	24.21	20.30
% non-English home language	1.65~	0.92	5.53~	2.88	37.28	31.75
% Disabilities	0.76~	0.43	2.54~	1.36	18.36	15.82
% African American	-1.56	0.87	-5.23	2.84	26.00	31.23
% Asian	-0.77	0.47	-2.60	1.43	9.11	11.71
% Hispanic	3.87**	1.23	13.02**	3.87	34.94	21.92
% White	-1.33	1.28	-4.48	4.07	26.17	30.65
% Female	-0.09	0.22	-0.31	0.70	48.67	48.98
Student performance						
ELA % Advanced or Proficient (Grade 3-5 tests)	-1.00	0.88	-3.36	2.71	49.36	52.72
Math % Advanced or Proficient (Grade 3-5 tests)	-0.93	0.86	-3.13	2.62	48.10	51.23
Teacher and school characteristics						
% Licensed teachers	1.80	1.20	6.12	3.85	89.00	82.88
Student-teacher ratio	-0.01	0.16	-0.04	0.50	14.04	14.08
% Exemplary teachers	1.50~	0.79	5.23~	2.59	21.15	15.92
% Proficient teachers	-1.27	0.92	-4.41	3.03	74.35	78.76
% Teacher retention	1.07	0.95	3.54	2.97	74.38	70.84
Stability	-0.10	0.26	-0.33	0.83	91.25	91.58
Class Size	-0.07	0.17	-0.22	0.55	20.01	20.23

Note. This table presents school-level measures from the school each student was enrolled for the longest period of time in fifth grade. Across student characteristic and student performance variables, data were missing for 13% of treatment students and 21% of control students. Data were missing for 26% of treatment students and 35% of control students on teacher rating measures, 29% of treatment students and 35% of control students on student-teacher ratio, 28% of treatment students and 32% of control students on percent licensed teachers and 45% of students on teacher retention measures. High rates of missingness of teacher and school characteristics are due to missing data at the school-level in year prior to 2017-18. Statistical significance levels are indicated as: ~ $p<0.10$; *$p<0.05$; **$p<0.01$;***$p<0.001$. ELA = English Language Arts

Table 4: CACE Estimates of 7th Grade Context Differences Between Lottery Winners and Control Group

	ITT	Standard Error	CACE	Standard Error	Lottery winner compliers mean	Control group compliers mean
Student background characteristics						
% High Needs	1.24	1.04	4.30	3.46	55.44	51.14
% English learners	1.13	0.72	3.92	2.43	15.39	11.47
% non-English home language	1.95*	0.89	6.74*	2.86	37.38	30.64
% Disabilities	0.11	0.48	0.38	1.59	15.73	15.35
% African American	-0.54	0.64	-1.88	2.07	25.72	27.60
% Asian	0.05	0.52	0.19	1.70	12.15	11.96
% Hispanic	3.07***	0.86	10.62***	3.10	30.54	19.92
% White	-2.63*	1.08	-9.09*	3.70	27.99	37.08
% Female	0.29	0.20	1.00	0.64	50.63	49.63
Student performance						
ELA % Advanced or Proficient (Grade 6-8 tests)	-0.21	1.24	-0.72	4.06	52.10	52.82
Math % Advanced or Proficient (Grade 6-8 tests)	-0.63	1.26	-2.19	4.18	50.06	52.25
Teacher and school characteristics						
% Licensed teachers	-0.31	0.58	-1.10	1.98	91.62	92.72
Student-teacher ratio	0.24	0.23	0.83	0.71	14.32	13.49
% Exemplary teachers	1.86*	0.83	6.19*	2.61	22.92	16.73
% Proficient teachers	-1.85*	0.89	-6.13*	2.80	72.74	78.87
% T retention	-0.57	0.67	-1.90	2.16	79.51	81.41
Stability	-0.11	0.30	-0.39	0.99	91.91	92.30
Class Size	0.10	0.27	0.34	0.87	20.89	20.55

Note. This table presents school-level measures from the school each student was enrolled for the longest period of time in eighth grade. Across variables, data were missing for 17% of treatment students and –22% of control students. Data were missing for 29% of treatment students and 35% of control students on teacher rating measures, 27% of treatment students and 31% of control students on student-teacher ratio, 27% of treatment students and 30% of control students on percent licensed teachers and 40% of students on teacher retention measures. High rates of missingness of teacher and school characteristics are due to missing public data at the school-level in year prior to 2017-18. Statistical significance levels are indicated as: ~*p*<0.10; * *p* <0.05; ** *p* <0.01;*** *p* <0.001. ELA = English Language Arts.

Table 5: Algebra sensitivity analysis

	By end of 8th grade				By end of follow-up year 9			
	ITT	Standard Error	CACE	Standard Error	ITT	Standard Error	CACE	Standard Error
Took Algebra I	6.55*	2.44	22.64**	7.98	7.17**	2.37	24.65**	7.78
Passed Algebra I	6.52*	2.55	22.57**	8.06	7.00*	2.56	24.22**	8.18

Note: Statistical significance levels are indicated as: $*p<0.05$; $**p<0.01$.

Table 6: Attendance sensitivity analysis

	Attendance Rate (%)		Number of days attended		N days with unexcused absences	
	CACE	Standard Error	CACE	Standard Error	CACE	Standard Error
4th grade	-0.30	0.75	-2.98	2.13	1.48	0.99
5th grade	-1.52*	0.67	-1.03	2.67	1.61~	0.94
6th grade	-1.52*	0.76	-2.73	1.90	2.23*	0.87
7th grade	-1.66**	0.62	-1.50	2.63	2.59*	1.31
8th grade	-0.25	0.98	-4.42	2.75	2.92*	1.20

Note: Statistical significance levels are indicated as: ~p<0.10; *p<0.05; **p<0.01

Table 7: Missing covariate sensitivity analysis

	Listwise deletion		Multiple imputation	
	ITT	*SE*	ITT	*SE*
Enrolled in BPS (%)				
BPS Enrolled 4-5	8.47**	2.60	9.75***	2.68
BPS Enrolled 6-8	5.14~	2.65	5.95*	2.63
Enrolled in a Charter School (%)				
Charter Enrolled 4-5	-3.63**	1.22	-3.60**	1.21
Charter Enrolled 6-8	-2.71	1.84	-1.90	1.80
Exam Schools (%)				
Took exam school test in 6th-8th	4.19~	2.35	5.39*	2.36
Admitted to exam school 7th-8th	3.04	2.52	4.05	2.59
Enrolled in exam school 7th-8th	1.44	2.22	2.31	2.26
School Progress/engagement				
Ever SPED 4-8	1.33	1.77	1.38	1.75
Ever retained 3-7	-0.10	0.64	-0.17	0.66
Average attendance rate (4-8)	-0.15	0.15	-0.11	0.21
Ever suspended (4-8)	-1.60	1.32	-1.77	1.32
Algebra I Outcomes				
Took Algebra I by 8th grade	6.55**	2.44	6.45**	2.43
Passed Algebra I by 8th grade	6.52*	2.55	6.50*	2.54
State Reading Test Scores				
4th grade	0.02	0.04	0.01	0.04
5th grade	0.03	0.04	0.03	0.04
6th grade	0.05	0.05	0.05	0.05
7th grade	0.05	0.04	0.05	0.04
8th grade	0.04	0.05	0.04	0.05
State Math Test Scores				
4th grade	-0.00	0.06	-0.00	0.06
5th grade	0.03	0.04	0.02	0.04
6th grade	0.05	0.04	0.05	0.04
7th grade	0.07*	0.03	0.07*	0.03
8th grade	0.01	0.05	0.01	0.05

Note: Outcomes were missing data as follows: 0% for BPS enrollment outcomes; 11-19% for charter school outcomes across variables; 0% for exam school (unconditional) outcomes; 3-14% for school progress/engagement across variables; 17-24% for algebra outcomes across variables; 15-25% for test scores in 4th-7th grades; and 42-46% for 8th grade test scores (due to COVID-19 closures when the final cohort was mostly in 8th grade). Statistical significance levels are indicated as: ~p<0.10; * $p < .05$; ** $p < .01$; *** p<.001. ITT = intent-to-treat; BPS = Boston Public Schools; SPED = special education.

Table 8: Test score impacts robustness to standardized test score decisions

	Score 1: First test taken in requisite grade		Score 2: Normal grade progression students only		Score 3: Tests included until student off track		Score 4: Dropped retained/skipped a grade students from all years	
	ITT	CACE	ITT	CACE	ITT	CACE	ITT	CACE
State reading test scores								
4th grade	0.02	0.05	0.01	0.03	0.01	0.03	0.00	0.01
	(0.04)	(0.13)	(0.04)	(0.12)	(0.04)	(0.12)	(0.04)	(0.13)
5th grade	0.03	0.10	0.02	0.06	0.02	0.06	0.01	0.03
	(0.04)	(.13)	(0.04)	(0.13)	(0.04)	(0.13)	(0.04)	(0.13)
6th grade	0.05	0.17	0.04	0.15	0.04	0.15	0.05	0.17
	(0.05)	(0.15)	(0.5)	(0.16)	(0.05)	(0.16)	(0.05)	(0.16)
7th grade	0.05	0.17	0.03	0.11	0.03	0.11	0.04	0.14
	(0.04)	(0.13)	(0.04)	(0.14)	(0.04)	(0.14)	(0.04)	(0.14)
8th grade	0.04	0.14	0.04	0.12	0.03	0.08	0.03	0.09
	(0.05)	(0.15)	(0.06)	(0.16)	(0.06)	(0.16)	(0.05)	(0.15)
State math test scores								
4th grade	-0.00	-0.00	-0.00	-0.01	-0.00	-0.01	-0.01	-0.03
	(0.06)	(0.17)	(0.05)	(0.16)	(0.05)	(0.16)	(0.05)	(0.17)
5th grade	0.03	0.09	0.03	0.09	0.03	0.09	0.03	0.11
	(0.04)	(0.12)	(0.04)	(0.11)	(0.04)	(0.11)	(0.04)	(0.12)
6th grade	0.05	0.18	0.05	0.18	0.05	0.18	0.07~	0.23~
	(0.04)	(0.14)	(0.04)	(0.13)	(0.04)	(0.13)	(0.04)	(0.13)
7th grade	0.07*	0.24*	0.07*	0.25*	0.07*	0.25*	0.07*	0.24*
	(0.03)	(0.12)	(0.03)	(0.12)	(0.03)	(0.12)	(0.04)	(0.12)
8th grade	0.01	0.04	0.00	0.01	-0.00	-0.01	0.03	0.09
	(0.05)	(0.14)	(0.05)	(0.14)	(0.05)	(0.13)	(0.05)	(0.14)

Note: Statistical significance levels are indicated as: ~p<0.10; * $p < .05$; ** $p < .01$;*** $p < .001$. Standard errors are shown in parentheses.

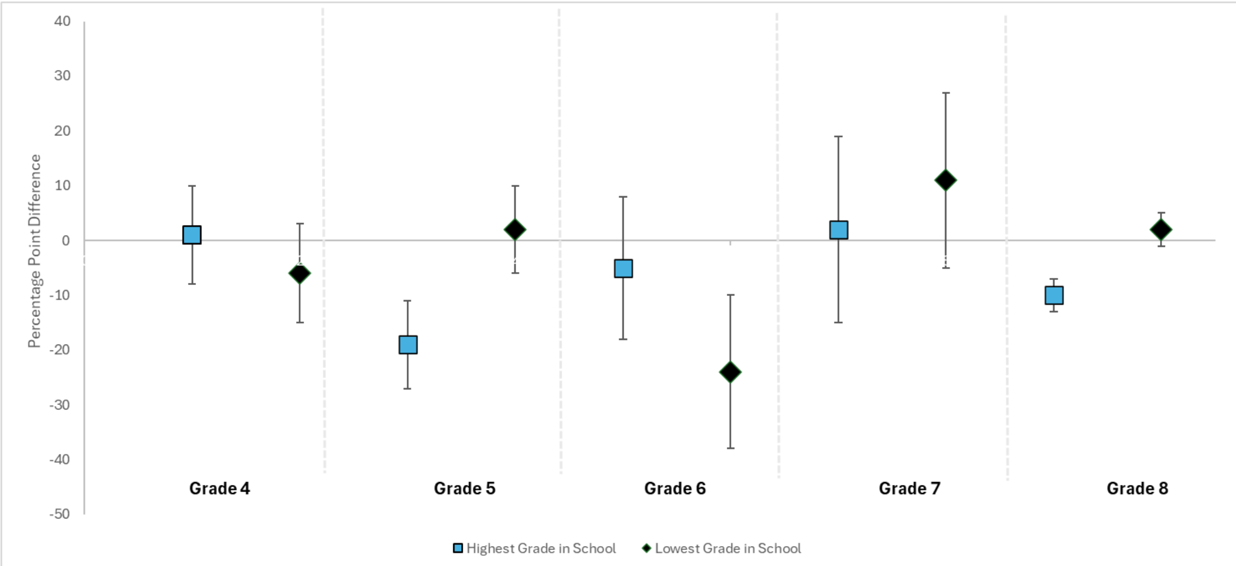
+Data are missing for most of the last cohort in 8th grade due to the cancellation of state testing in spring 2020.

Table 9. Replication Weiland et al. (2020) Table 3: Replication of First Choice Lottery Sample ITT Impacts and CACE

Variable	ITT	ITT (SE)	First Stage	First Stage (SE)	CACE	Lottery Winner Compliers Mean	Control Group Compliers Mean
Enrolled in BPS (%)							
Pre-K	29.13***	3.16	--	--	--	100	0
Kindergarten	7.12***	1.74	0.29***	0.03	24.44***	91.23	66.79
First grade	10.27***	2.25	0.29***	0.03	35.25***	85.86	50.61
Second grade	8.62***	2.53	0.29***	0.03	29.6***	79.15	49.55
Third grade	7.67***	2.80	0.29***	0.03	26.34***	75.75	49.41
Enrolled K-3	8.30***	2.90	0.29***	0.03	28.48***	73.58	45.1
Ever enrolled	6.02***	1.42	0.29***	0.03	20.65***	100	79.35
Retained in grade (%)							
Retained in kindergarten	0.02	0.28	0.30***	0.03	0.08	0.21	0.13
Retained in first grade	0.01	0.68	0.30***	0.03	0.04	1.51	1.47
Retained in second grade	1.22	0.84	0.30***	0.03	4.08	4.28	0.2
Ever retained	1.16	0.98	0.29***	0.03	3.97	7.53	3.56
Special education classification (%)							
Kindergarten	-0.32	1.32	0.30***	0.03	-1.06	7.79	8.85
First grade	0.19	1.49	0.30***	0.03	0.62	10.92	10.3
Second grade	1.42	1.72	0.30***	0.03	4.73	13.52	8.79
Third grade	-0.2	1.75	0.30***	0.03	-0.68	16.01	16.69
Ever SPED	0.41	1.62	0.29***	0.03	1.40	17.34	15.94
MCAS and PARCC							
Math	-0.06	0.04	0.29***	0.03	-0.19	0.38	0.57
English language arts	0.00	0.04	0.30***	0.03	0.01	0.43	0.42

Note: Statistical significance levels are indicated as: *$p < .05$; ** $p < .01$; *** $p < .001$. ITT = intent-to-treat; CACE = complier average causal effect; BPS = Boston Public Schools; MCAS = Massachusetts Comprehensive Assessment System; PARCC = Partnership for Assessment of Readiness for College and Careers.

Figure 1: CACE Estimates of Impact of Being in the Youngest Grade or Oldest Grade Offered by School Between Lottery Winners and Control Group



Note: This figure displays the CACE estimates of the impact of enrolling Boston Pre-K on the likelihood of being enrolled in the highest grade offered by students' school (square marker) and on the likelihood of being enrolled in the lowest grade offered by their school in each follow-up year. Black bars represent 95% confidence intervals.

Appendix C: Details on the full applicant sample and IPW work

Following other lottery-based studies (i.e., Abdulkadiroğlu et al., 2011) and building from Weiland et al. (2020), we explored the external validity of Boston Pre-K estimates. We did so because our lottery sample children represent only 24% of Boston Pre-K applicants in all. We present in this appendix additional details on this generalizability work.

Outcome data availability. In the full sample, as shown in Table 1, outcome data were 10-16 percentage points more likely to be missing for non-enrollees than enrollees (p<.001). Table 2 presents the background characteristics across groups (enrolled vs non-enrolled) for students with any non-missing standardized tests data from 4th to 8th grade. There were differences between enrollees and non-enrollees in terms of race/ethnicity (e.g., enrollees were 5pp less likely to be Hispanic, p<.001), eligibility for free/reduced lunch (e.g., enrollees were 6pp more likely to be eligible, p<0.001), age (e.g., enrollees were approximately 0.26 years older, p<0.001), country of origin (e.g., enrollees were 9pp more likely to report the U.S. as their country of origin, p<.001), and home language (e.g., enrollees were 2pp less likely to speak English as their home language, p<.10). The F value from a joint F-test in the full sample was statistically significant (p<.001). Data availability and student characteristic differences between enrollee and non-enrollee sample underscore cautions in the main text about interpreting estimates from this sample as *associations* only, meant to give some information about generalizability of the lottery-based ITT/CACE estimates.

Covariate balance in IPW work. Table 3 displays differences in covariates in the ITT replication and full samples before and after IPW reweighting. Re-weighting resulted in smaller differences between relevant groups in both samples except for students who identified as "Other" for race/ethnicity in the full sample, where the reweighted difference was marginally (0.01) higher.

Table 1: Percent of missing data on key outcomes in the full applicant sample

	Enrollee	Non-enrollee	Difference	*P*-value
Enrolled in a Charter School (%)				
Charter Enrolled 4-5	9.97	25.16	-15.20***	0.000
Charter Enrolled 6-8	11.16	24.96	-13.80***	0.000
Charter Ever Enrolled 4-8	8.09	21.42	-13.32***	0.000
School Progress/Engagement				
Ever SPED 4-8	1.03	12.79	-11.76***	0.000
Ever retained 3-7	6.14	18.76	-12.62***	0.000
Average attendance rate (4-8)	8.09	21.44	-13.35***	0.000
Ever suspended (4-8)	8.09	21.42	-13.32***	0.000
Algebra I Outcomes				
Took Algebra I by 8th grade	15.30	28.21	-12.92***	0.000
Passed Alegra I by 8th grade	16.81	29.74	-12.93***	0.000
Standardized Test Reading				
4th grade	14.04	30.13	-16.09***	0.000
5th grade	14.80	30.33	-15.53***	0.000
6th grade	16.67	31.75	-15.09***	0.000
7th grade	18.43	32.52	-14.09***	0.000
8th grade	39.61	50.95	-11.34***	0.000
Standardized Test Math				
4th grade	13.86	30.21	-16.35***	0.000
5th grade	14.94	30.81	-15.87***	0.000
6th grade	16.79	32.19	-15.39***	0.000
7th grade	18.82	32.76	-13.94***	0.000
8th grade	39.81	50.99	-11.17***	0.000

Note: BPS enrollment, BPS persistence, and exam school variables are not shown in the table because there were no missing data for these outcomes. Statistical significance levels are indicated as: ***p<.001.; SPED = special education.

Table 2: Balance on baseline characteristics for students with any test score data from grades 4-8 in the full applicant sample

	Enrollee	Non-enrollee	Estimated association	*P*-value
Race/ethnicity (%)				
Hispanic	42.37	47.68	-5.31***	0.000
Black	28.72	25.62	3.10***	0.000
White	16.51	13.53	2.98***	0.000
Asian	9.32	4.96	4.35***	0.000
Other	3.08	2.70	0.38	0.270
Male (%)	51.00	49.34	1.66	0.101
Eligible for free/reduced lunch (%)	70.69	64.84	5.85***	0.000
Age	4.53	4.27	0.26***	0.000
Country of origin USA (%)	95.05	86.26	8.80***	0.000
Home language (%)				
English	47.59	49.39	-1.80~	0.066
Spanish	30.83	27.22	3.60***	0.000
Other	21.59	17.78	3.81***	0.000
N children	7,116	3,723		

Note: For children with non-missing math and ELA test scores, there was no missing data on race/ethnicity, sex, or age. There was a small amount of missing data on other variables: 126 students were missing free/reduced lunch information, 15 were missing country of origin information, and 273 were missing home language information. Values for enrollees are the simple means for each requisite group. Values for the difference between enrollees and non-enrollees were obtained from a regression of a given baseline characteristic and a vector of neighborhood level fixed effects on the enrollment indicator. The coefficient on the enrollment indicator equals the difference in the mean baseline characteristic between enrollees and non-enrollees. The value for non-enrollees equals the corresponding value for enrollees minus the estimated difference between enrollees and non-enrollees, respectively. A two-tailed t-test was applied to the estimated difference. Statistical significance levels are indicated as: ~p<.10; ***p<.001. An F-test was used to assess the statistical significance of the overall difference between enrollees and non-enrollees reflected by the full set of baseline characteristics in the table. The resulting F-value was statistically significant.

Table 3: Example covariate differences between treatment and control members in the lottery replication sample and the full sample before and after IPW

	ITT Replication Sample		Full Sample	
	Unweighted	Reweighted	Unweighted	Reweighted
Race/ethnicity (%)				
Hispanic	-11.33	0.37	-14.92	0.24
Black	8.40	-0.78	3.21	-0.22
White	-4.58	-0.52	3.91	-0.26
Asian	15.60	2.31	16.97	0.86
Other	-2.09	-1.51	0.89	-0.90
Male (%)	1.14	-0.07	-3.40	-0.09
Eligible for free/reduced lunch (%)	25.84	0.07	8.31	0.26
Age	3.95	-0.18	3.71	-0.01
Country of origin USA (%)	-1.69	0.69	1.48	0.25
Home language (%)				
English	-16.50	-0.64	-14.26	-0.23
Spanish	9.66	-0.84	7.52	-0.32
Other	9.69	1.73	9.23	-0.65
N children				
Treatment	6,820	4,619	7,590	5,793
Control	2,438	4,638	4,008	5,805

Notes: Results shown for the “BPS Enrolled 4-5” outcome. Neighborhood fixed effects corresponding to the neighborhood a child lived in during his or her preschool-eligible year (as proxied by a fixed effect for the public elementary closest to the child’s home) were also included in IPW models but are not shown here for parsimony. Full results and results for other outcomes available upon request.