Pre-K in the Public Schools: Evidence from Within All States[1]

Timothy J. Bartik and Brad Hershbein[2]

November 2015

ABSTRACT

Within the past decade, state-funded pre-Kindergarten has roughly doubled in its coverage of 4-year-olds, and further large-scale expansion of pre-K programs, with state, local or federal funding, continues to be debated. Although research has shown that pre-K can increase test scores and dramatically improve life outcomes, at least for some programs at some places and times, existing studies have generally focused on small or state-specific programs that may not sufficiently capture program heterogeneity and thus may not generalize to other areas or programs. In this paper, we draw upon multiple data sources to exploit variation in enrollment in public pre-Kindergarten programs across time and place to examine the effect of these programs on standardized test scores and other academic outcomes. Our data cover the last two decades, span nearly all states, and allow for intrastate variation in pre-K. We investigate both program-level heterogeneity as well as heterogeneous impacts on different types of students and schools. To our knowledge, this set of analyses is the first to provide national-level estimates of effects of public pre-Kindergarten access on academic outcomes for different types of students and schools.

Keywords: pre-K, early childhood education, NAEP, test scores, sleeper effects

JEL Codes: H75, I21, I24, I28

<u> PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE OR CIRCULATE WITHOUT AUTHORS' PERMISSION **</u>**

[1] We thank the Russell Sage Foundation for its generous support of this project. Wei-Jang Huang provided invaluable research assistance. All errors are our own.

[2] Bartik, Hershbein: W.E. Upjohn Institute for Employment Research. Address: W.E. Upjohn Institute, 300 S. Westnedge Ave., Kalamazoo, MI 49007.

I. Introduction

In recent years researchers have documented stagnant educational achievement gaps across races and rising inequality between children from poor and richer families. Reardon (2011), for example, has shown that black students still lag behind white students by roughly 1.5 years of learning on standardized tests. The gaps in the same test scores between students from the 90th percentile of family income and those from the 10th percentile of family income have grown by over 30 percent over the past 30 years. These differences in achievement correspond to widening disparities in educational attainment. Students from the bottom quartile of the income distribution nearly doubled their likelihood of completing college, from 5 percent to 9 percent, for cohorts born in the early 1960s and those born twenty years later; however, students from the top income quartile saw a much larger increase, from 36 percent to 54 percent (Bailey and Dynarski 2011). Even very recent evidence from the Nation's Report Card, released in October 2015, show that gaps by race and family income have barely budged in the last decade. As economic opportunity and social mobility rest, in large part, on educational achievement and attainment, policymakers have expressed increased demand for policy levers to ameliorate these racial and socioeconomic inequalities.

One such policy lever is early childhood education, specifically publicly-funded pre-kindergarten (pre-K) programs. These programs aim to provide skills to young children so that they are better prepared to learn once they enter schooling with a more formal curriculum. Public pre-K acts as an alternative to other early childhood programs. Although private preschool programs have long existed, they can be expensive for families of modest means. Federal programs, notably Head Start, are often over-subscribed and tend not to reach many of the near-poor, who may still qualify for other government assistance programs such as Medicaid, SNAP (food stamps), or federal lunch assistance. The perceived simplicity and perhaps greater universality of public pre-K, especially if provided through the public schools, has led advocates to call for their expansion. Mayor Bill de Blasio of New York City, for example, has implemented a universal pre-K program for some 65,000 of his city's 4-year-olds (Siegel 2015). More generally, state-funded pre-K programs have already expanded from covering 14 percent of all 4-year-olds in 2001–2002 to 29 percent in 2013–2014 (Barnett et al., 2014). Are these programs effective in reducing inequities in education among young students?

Substantial volumes of research show that pre-K *can* be effective in raising test scores in early grades, as well as boosting educational attainment, earnings, and health, and reducing crime. However, most studies are of small programs, or of ones that took place several decades ago, or of those that targeted very narrow groups of students. The most celebrated programs, Perry and Abecedarian, were all three. Some more recent studies look at newer pre-K programs introduced on a statewide scale, but only in one or two states. Very few studies examine pre-K programs throughout the country, and it is unclear whether the findings from local or even state programs generalize. Additionally, many of the specific programs studied may be of unusually high "quality" and may not reflect typical pre-K programs, as they have been implemented and as they may continue to be implemented. Furthermore, studies of statewide programs may not adequately control for other factors that could influence educational achievement and impact estimates are often relatively noisy, leading to uncertainty about their true effect.

In this paper we perform the first national analysis of public pre-K participation on standardized test scores, special education assignment, and grade retention using substate variation. We match detailed microdata of the National Assessment of Educational Progress (NAEP), the Nation's Report Card, to public pre-K enrollment at the state, district, and school level for different types of students and schools. For the 4th and 8th graders taking the NAEP, we use the Common Core of Data from the National Center for Education Statistics to estimate their likelihood of being enrolled in pre-K five (or nine) years ago. Our data stretch from pre-K enrollments in the early 1990s (4th grade outcomes in the mid-to-late 1990s) through pre-K enrollments in 2008 (4th grade outcomes in 2013), offering substantial variation in public pre-K over time and space.

To identify the impact of pre-K on student outcomes, we adopt a two-stage augmented differences-in-differences methodology. The first stage uses student-level data in NAEP to calculate means at the geography-year cell net of individual student characteristics. The second stage takes these collapsed means and implements differences-in-differences controlling for geography and time fixed effects, and sometimes higher-level interactions. The extent of pre-K variation allows for more precise estimates than most previous studies, although it comes at the expense of program specificity. That is, instead of estimating the effect of a specific pre-K program on later outcomes, we effectively estimate the "average" effect of pre-K diffusion through public schools on both academic and non-academic outcomes. The data allow us to estimate effects for students overall as well as for different groups of students (or schools or districts), stratified by race, income, and other characteristics.

Although there is some evidence for a reduction in the likelihood of receiving special education services, in general we find precisely estimated zeros across a variety of outcomes, student groups, and geographies. In many cases, we can rule out effects as small as a 1 percentile test score improvement when moving from a zero share of students enrolled in pre-K to a 100-percent share of students enrolled in pre-K. Similarly, we can often rule out impacts as small as a 1.5 percentage point reduction in grade retention.

While our measure of pre-K is uncommon (but not unique) and cannot be perfectly matched to individual students, we do not believe our null findings are due to measurement error in this variable, as our standard errors are small and we validate our pre-K measure against other sources. We also do not believe our findings are due to endogeneity of changes in pre-K access. We control for a given jurisdiction's fixed effects on student success, which should control for the main form of endogeneity, and we also control for higher-order geographies (states) interacted with year of observation to control for nonlinear policies. It is conceivable that our results could be due to unobserved trends in student outcomes that are extremely persistent over time and that lead to a jurisdiction both choosing to invest more in pre-K and then experiencing deteriorating student outcomes due to pre-existing trends. However, we view this scenario as unlikely. Rather, we argue our results are consistent with the broader literature finding fadeout of test score effects in middle grades, which occur even among high-quality programs that have been shown to have large short-term and long-term (but not middle-term) impacts. Furthermore, our analysis averages across all qualities of pre-K programs in public schools, so middle-term impacts should be even smaller than in high-quality programs.

Our study, like any other, is not without its weaknesses. We cannot yet fully account for the counterfactual that students may be enrolled in other early childhood education (e.g., Head Start, private centers).[3] Additionally, we have not yet ruled out that typical pre-K programs may have long-term "sleeper" effects; these could arise due to pre-K affecting skills that are unobserved in our data but that affect later outcomes such as high school graduation or the propensity to commit crime (Smith 2015). Nonetheless, our findings suggest that the expansion of pre-K, at the quality levels at which it has typically been implemented in public schools over the last two decades, has not had an effect on middle-term outcomes on average, even for relatively disadvantaged students. Our findings do **not** imply that there are no pre-K programs (perhaps of higher quality) that had an effect on middle-term outcomes, nor do they imply that typical pre-K programs have no effect on longer-term outcomes.

The remainder of this paper is organized as follows. The next section outlines the conceptual and methodological challenges in estimating the impact of pre-K on social, educational, and economic outcomes. In the context of these challenges, we review and interpret findings from the large and growing pre-K research literature. We then describe our empirical approach, and its advantages and

[3] Indeed, our findings could possibly be biased by the existence of Head Start as an option, just as it appears some of the Head Start research is biased by the existence of state pre-K as an option (Kline and Walters 2015). We discuss a possible strategy to reveal this counterfactual in the conclusion.

shortcomings relative to previous studies. The following section presents our results and places them in the context of existing literature. We conclude with next steps.

II. The Research Literature on Pre-K: Implications for This Study

In this section, we briefly summary the large research literature on pre-K effects. Our summary focuses on the research findings and limitations that are most relevant to our current study. Appendix Table A1 provides a more detailed listing of results from most of the prominent pre-K studies, including estimated pre-K effects at the short-term, medium-term, and long-term. The present text section relies on this table in stating pre-K effects; the appendix table provides references to specific studies, and how we report each study's results.

Six aspects of the pre-K research literature seem most relevant to the current study.

(1) <u>Select pre-K programs versus average pre-K programs.</u> Several experimental studies, and many good non-experimental studies, find large long-run and short-run effects of pre-K on student outcomes. However, these studies by necessity are limited to selected programs—often higher quality programs—and may not apply to average state and local public pre-K programs, which is what we examine in this study. Two classic experimental studies from the 1960s and 1970s, Perry Preschool and the Abecedarian program, have found large effects on former participants' outcomes, both in the short-run and long-run. Long-run effects on earnings, for example, are 19 percent in Perry, and 26 percent in Abecedarian. Short-term effects (e.g., at the end of pre-K or beginning of kindergarten) in both studies included increasing test scores by almost 20 percentiles.[4] However, Perry Preschool and the Abecedarian program are far more intense than usual pre-K programs, with Perry costing over $20,000 per student (in today's dollars) and Abecedarian over $85,000.

Other studies, of pre-K programs that are somewhat closer to typical programs, have also found short-run and long-run effects, of perhaps one-third to two-thirds the effects found for Perry and Abecedarian. A quasi-experimental study of Head Start estimated adult outcome effects that predict a Head Start earnings boost of 11 percent, and short-run test score effects of 5 percentiles (Deming 2009). Studies of the Chicago Child-Parent Study estimate long-run earnings effects of 8 percent, and short-run test score gains of 11 percentiles. Summarizing many similar studies, meta-analyses of the pre-K literature find short-run test score effects that average 9 to 14 percentiles (Camilli et al. 2010 and

[4] We state test score effects in percentiles because evidence from Chetty et al. (2011) suggests tests scores measured in percentile units are linearly related to dollars of adult earnings across most of the income distribution. Appendix Table 1 also states effects in the more usual effect size units.

Duncan and Magnuson 2013).[5] However, even though these studied programs are often closer to typical state pre-K programs, most are probably higher quality than the typical state or local pre-K program. For example, the oft-cited pre-K programs in Chicago and Tulsa both spent over $5,000 per student for a half-day pre-K program for one school year, which is much more than most state and local pre-K programs spend per half-day. Another pre-K program that has received much attention and favorable evaluations is run by Boston Public Schools, which for a full-day program, spends over $15,000 per school-year per student.

An important issue is to what extent these results from very intense pre-K programs (Perry, Abecedarian) and less intense but still high-cost pre-K programs (many of the other studies) are matched by those from average state and local pre-K programs. We focus on these latter programs in the current study.

(2) Medium-term fading of pre-K test score effects. A second relevant aspect of the pre-K literature is the common finding of extensive, but not total, fading of test score effects between kindergarten and middle grades (3rd through 8th grades). In the meta-analyses of pre-K studies, test score effects decline by one-half to two-thirds over this horizon, with average medium-run test score effects of 4 to 5 percentiles. Some studies find more complete fading: Chicago CPC, Head Start, Tennessee, and Perry (but not in Abecedarian, which finds more modest fading). Still, the general pattern of the research results would lead one to expect that if the typical state and local pre-K program has comparable effects to the programs previously selected for study, we should be able to detect their effects in 4th grade test score outcomes, which are among the outcomes considered in the current study. As we will see, our data allow sufficient precision to easily detect test score effects in some specifications that are much smaller than 4 percentiles.

(3) Sleeper effects of pre-K re-emerging in the long-term. A third relevant aspect of pre-K studies, and indeed of early childhood programs of any type, is that the short-run to medium-run fading of test score effects is often followed by recovery of large adult effects later. This pattern is pronounced in the Perry program, the Chicago CPC program, Deming's (2009) study of the Head Start program, and Chetty et al.'s (2011) study of the effects of higher "kindergarten class quality." These re-emerging "sleeper effects" have sometimes been argued to be due to program effects on soft skills (Heckman et

[5] The age-based regression discontinuity design studies in Appendix Table 1 in Tulsa, Boston, and Tennessee find somewhat larger effects. This may reflect that the regression discontinuity studies, which compare pre-K entrants with pre-K graduates/kindergarten entrants who are on either side of the age cut-off, are comparing pre-K graduates with a control group that is further away in age from entering kindergarten, and therefore less likely to have attended pre-K the previous year than is true of comparison groups in other pre-K studies. The counterfactual in RDD pre-K studies involves students who have less education.

al. 2013, Heckman 2015), which are known to be difficult to measure (Duckworth and Yeager 2015). If soft skills are important to long-term effects, then it is important to try to measure pre-K effects on outcomes that might be more correlated with soft skills than is sometimes argued to be the case for standardized test scores. We try to do this in the current study by looking at pre-K's effects on grade retention and assignment to special education status at both 4th and 8th grade. As this project progresses, we plan to further push this line of inquiry by looking at longer-term behavioral outcomes, such as high school graduation.

(4) Geographic access studies of pre-K. There is a conceptual difference between studies that focus on variation in individual-level access to pre-K in a specific area and those that look at variation in average access to pre-K across geographic areas or other groups. These latter, geographic studies often seem to find surprisingly large test score effects in the medium-run.[6] For example, Cascio and Schanzenbach's (2013) study comparing pre-K in Georgia and Oklahoma with other states finds 4th grade test score effects of 14 percentiles. Ladd et al.'s (2014) study comparing counties with different pre-K access in North Carolina finds third grade test scores of 20 to 25 percentiles. As Ladd et al. point out, if "there were no spillover effects of the program to other children, the test score impacts would be unrealistically large." But such spillover effects make sense, given evidence of the importance of peer spillovers found in Hanushek et al. (2003) and Hoxby (2000), as well as direct evidence for positive spillovers in kindergarten from more students in pre-K found in Neidell and Waldfogel (2010). These very high medium-run effects found in some geographic access studies provide another support for the ability of the current study to find pre-K test score effects, if such test score effects exist for the average state and local pre-K program.

(5) Difficulty of measuring pre-K quality. A fifth relevant aspect of the pre-K research literature is that everyone agrees that quality is important, but the empirical evidence suggests that we do not know how to measure quality in a way that shows consistent effects on child outcomes. There are often only modest or inconsistent relationships between existing structural measures of pre-K quality (teacher credentials, class size, written curriculum, classroom physical features) and student learning in pre-K (Zaslow et al. 2010; Sabol et al. 2013; Bartik 2011, pp. 135–140). Furthermore, observational measures of pre-K quality (e.g., having trained outside observers try to objectively rate the average quality of

[6] Although these large medium-run effects occur for Cascio and Schanzenbach (2013) and Ladd et al. (2014), medium-run effects are not large in Fitzpatrick (2008) or Rosinsky (2014). Rosinsky's results are quite sensitive to the inclusion of particular states. Also, Cascio and Schanzenbach's results and Fitzpatrick's results are relatively imprecise due to only having one or two treatment states, with statistical significance sensitive to how the standard errors are treated. When the standard errors are adjusted as suggested by Conley and Taber (2011), the standard errors become large enough that one cannot reject zero pre-K effects or very large pre-K effects.

teacher-student interactions) also are not always strongly and systematically correlated with higher pre-K learning. Some studies have found some modest positive correlations between CLASS (Classroom Assessment Scoring System) quality ratings and student learning (Keys et al. 2013), but not for other observational rating systems. However, other studies have found that higher CLASS ratings do not always predict better student outcomes (Burchinal, Kainz, and Cai 2011; Weiland et al. 2013). Overall, this pre-K research suggests that "currently available quality measures may not be adequate to the research tasks being undertaken" (Keys et al. 2013). Because of the recognized importance of quality, the current study will attempt to see whether pre-K's effects vary with current quality measures, but the past research does not give great confidence that conventional quality measures will help much.

(6) Importance of the counterfactual. A sixth aspect of the research literature is that pre-K's net impact can vary greatly depending upon what constitutes the counterfactual to a particular pre-K program. This has recently been shown to make a major difference in interpreting the results of the Head Start experiment, in which almost half of the randomly assigned control group attended some other early childhood program. Two recent papers show that Head Start's effects relative to a counterfactual of no preschool are about 60 percent greater than Head Start's net effect relative to a counterfactual that includes considerable preschool enrollment (Feller et al. 2014; Kline and Walters 2015).[7] Another recent paper shows how the diffusion of the television show *Sesame Street* in the late 1960s and early 1970s essentially functioned as an early childhood education program and improved schooling outcomes, in part because few children at the time were exposed to educational programming before elementary school (Kearney and Levine 2015). The counterfactual is also an issue we will have to consider in interpreting the results of our current study. What is being compared is pre-K programs in the public schools versus pre-K programs offered elsewhere, paid for either by private funds or government funds. We hope in future work to better control for the other options available in different geographic areas.

III. Data and Methodology

A. Data

Our data come from two main sources: the National Assessment of Educational Progress (NAEP), also called the Nations' Report Card, and the Common Core of Data (CCD). Both datasets are

[7] As mentioned in a previous footnote, a different counterfactual may also help explain the generally greater short-term test score effects found in regression discontinuity studies of pre-K.

maintained by the U.S. Department of Education. We supplement these sources with population data and pre-K program data from the National Institute of Early Education Research (NIEER).

NAEP

The NAEP is a nationally representative standardized assessment of academic subjects for certain grades, and is the only uniformly administered test that is comparable across states and time.[8] The core subjects of mathematics and reading are currently tested biennially, in odd-numbered years, at grades 4, 8, and 12. Since 2003, every state has participated in the core NAEP tests, and the large sample sizes—approximately 3,000 students per state for each test administration in grades 4 and 8—are sufficient to allow for detailed analyses of student groups. Prior to 2003, the math and reading tests for grades 4 and 8 were administered less frequently, about every four years, with participation by most but not all states.

NAEP data at the state level are publicly available (https://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx) and have been used in previous analyses of the effect of pre-K programs on student achievement (Grissmer, Flanagan, Kawat, and Williamson 2000; Cascio and Schanzenbach 2013; Rosinsky 2014). We employ, however, the restricted-access microdata, available to qualified researchers via license with the Institute of Education Sciences of the Department of Education. These microdata not only contain a wealth of information about individual students taking the NAEP and characteristics of the schools they attend, they also contain school and district identifiers that allow the data to be matched longitudinally over time and to be linked to external sources, such as the Department of Education's near-census of public schools, the Common Core of Data.[9]

The NAEP data provide our main outcomes of interest: math test scores, reading test scores, assignment to special education (i.e., has an Individual Education Plan), and a measure of whether children are over-age for their grade. NAEP test scores are provided (and reported publicly) as a scale score; we use both this measure and also convert to a percentile score using the 2013 NAEP score distributions for each grade and subject.[10] The percentile conversion is done because prior research by

[8] For more information, see https://nces.ed.gov/nationsreportcard/.

[9] To our knowledge, Fitzpatrick (2008) is the only previous paper to use the NAEP microdata to examine the effect of pre-K. However, she focused on the implementation of Georgia's universal pre-K program and did not exploit within-state variation. Chingos (2015) demonstrates how the microdata can be used for a much richer set of controls to more accurately measure comparisons in performance across students.

[10] To minimize burden, individual students take only a portion of the full test and item response theory is used to statistically impute multiple plausible scale scores for each student. We follow the literature and average these plausible scale scores for each student. The scale scores are approximately normally distributed.

Chetty et al. (2011) has shown that percentile test scores are linearly related to adult earnings measured in dollars. At 4th grade, Chetty et al.'s research suggests that a 1 percentile increase in test scores is associated with an increase in adult earnings of about one-half of 1 percent of overall mean adult earnings. At 8th grade, a 1 percentile increase in test scores is associated with an increase in adult earnings of 0.8 percent. Because lifetime earnings for the average American have a present value of around $730,000, a 1 percentile increase in test scores as of 4th grade (8th grade) would be predicted to increase the present value of future earnings by about $3,650 ($5,840).[11] In addition, because previous research has found that pre-K programs may improve later life outcomes through its effect on socioemotional as well as academic skills (Heckman, Pinto, and Savelyev 2013), we also examine the assignment to special education and whether a student is above the modal age for his or her grade. These latter outcomes are more likely to capture learning difficulties that reflect nonacademic as well as academic deficiencies.

CCD

The Common Core of Data (CCD) annually provides detailed characteristics of individual schools and school districts (local education agencies), including enrollment by sex, grade, and ethnicity, the share of students eligible for free-and reduced price lunch[12], pupil-to-teacher ratios, type of locale, and others.[13] Of greatest utility for this paper, the CCD reports counts of pre-K enrollment within the public schools. This measure is not ideal, as it does not capture pre-K programs that are publicly funded but operate in centers outside the public schools. This measure also does not account for enrollment in private pre-K programs, which are in some cases publicly subsidized (Barnett and Hustedt 2011).[14]

Nonetheless, we believe that enrollments from the CCD offer the best measure of spatial and temporal variation in the diffusion of pre-K. Some evidence suggests that pre-K programs located in

[11] These calculations follow from taking Chetty et al.'s (2011) raw dollar effects, and converting to percentage effects at overall mean earnings. We implement a slight downwards adjustment by taking the ratio of the "leave-out mean" estimates for kindergarten entrants to the ordinary least squares estimates in Chetty et al.'s Appendix Table XIII. See endnote 12 on page 80 in Bartik (2014) for more details. We derive the present value calculations from average earnings by age in the 2012 American Community Survey, discounted back to age 4 at 3 percent. Future earnings are assumed to increase by 1.2 percent per year from 2012 on.

[12] The National School Lunch Program provides subsidized school lunches for students in families whose income falls below 185 percent of the federal poverty guidelines.

[13] Some of these characteristics are also reported in the NAEP itself, but they are missing for a non-trivial number of schools and districts. The CCD also allows district financial data, including spending per-pupil, to be matched to NAEP, and we plan to do so in the future.

[14] Head Start, a federal preschool program intended for low-income students, may operate in partnership with public and private schools as well as standalone centers. We do not attempt to disentangle the source of funds used to pay for pre-K in the CCD enrollments.

public schools may be of higher average quality and lead to better results (Magnuson, Ruhm, and Waldfogel 2007), possibly because of better funding, better coordination with school expectations, and fewer transitions for children. Additionally, whereas previous papers (Fitzpatrick 2008; Cascio and Schanzenbach 2013) focused on the rollout of a universal pre-K program in one or two states, essentially making the adoption of pre-K into a binary event, the CCD counts offer changes in the intensive margin of pre-K for fifty states and the District of Columbia. This alone would provide advantages in estimation relative to previous studies, which typically employ few effective treatment groups and thus can suffer problems of inference (see Donald and Lang 2007 and Conley and Taber 2011). Furthermore, the CCD allow us to examine pre-K enrollment *at the district or even school level*, as these can be matched to identifiers within the NAEP dataset, something that has not been possible in previous research.

NIEER

However, a drawback of using CCD enrollment is that it cannot distinguish the quality of pre-K programs or even whether they are half-day or full-day. That is, the CCD measure implicitly treats all variation in public pre-K across locations and time as equivalent, which the early childhood education literature has categorically rejected (see Minervino 2014 for a review). As we pointed out earlier, while research has emphasized the importance of quality in pre-K programs, the research has not clearly identified quality measures that are consistently linked to student outcomes. Some of the most widely used quality benchmarks for state-funded pre-K programs come from NIEER, which since the 2001–2002 school year has produced annual yearbooks describing the size and nature of each state's pre-K programs. As part of their systematic data collection, NIEER also measures ten binary indicators of quality thresholds for each state's pre-K program(s) as well as a continuous measure of state and local expenditures per pupil.[15] Although these are relatively crude proxies for quality, they are the only ones to our knowledge that are available for every state.[16]

[15] The ten indicators are: (1) lead teachers required to have a bachelor's degree, (2) lead teachers required to have a specialization in early childhood education, (3) assistant teachers are required to have a Child Development Associate credential, (4) teachers are required to have at least 15 hours of annual in-service training, (5) class size is limited to 20 students, (6) student-teacher ratios are limited to no more than 10:1, (7)the state has comprehensive and appropriate early learning standards, (8) the state requires at least one meal in the program, (9) the state requires vision/hearing/health screening and parent involvement for referrals, and (10) the state monitors pre-K programs through site visits. A few states have multiple, separately-run public pre-K programs. For analysis purposes, we use enrollment-weighted averages in these cases.

[16] For example, the Classroom Assessment Scoring System (CLASS), which measures classroom interaction of students and teachers and in some studies is correlated with pre-K students' learning (Sabol *et al*. 2013), is proprietary and requires programs to opt in to participate.

We use these measures to construct quality indices at the state-level, as described below. Even though the NIEER enrollment data have the advantage of capturing public pre-K enrollment outside of public schools and are considered to be of sufficient quality that the U.S Department of Education has incorporated them into its official statistics, we have chosen not to make them our main measure of pre-K intensity for three reasons. First, the data extend back only through the 2001–2002 school year, and when matched to 4th grade NAEP test scores five years later result in less temporal variation than the CCD. Second, the enrollment data are available only at the state level, unlike the district- and school-level rates possible with the CCD. Third, the NIEER data do not disentangle enrollment at programs provided in public schools and those provided in stand-alone centers, and this mix varies across states and within states over time.

Comparing Pre-K Data Sources

Because our choice of pre-K enrollment is uncommon (but not unprecedented) in the literature, we have examined how the CCD measure compares to both the NIEER measure and enrollment rates derived from the Census and the American Community Survey.

To construct rates in the first two sources, we convert the pre-K enrollment counts into either *population shares* or *population ratios*, depending on the level of analysis. At the state level, we divide the annual count of students enrolled in pre-K programs in public schools by the annual estimate of a state's 4-year-olds, as provided by the SEER program of the National Cancer Institute.[17] We do this for both the CCD and NIEER. Thus, these population shares represent the fraction of a state's 4-year-olds enrolled in a public pre-K program in a given year. At the district and school levels, however, there is no reliable and consistent source for the annual count of 4-year-olds. We therefore construct a population ratio with the CCD data by dividing the count of pre-K enrollment by the count of first-grade enrollment at the same district or school in that year.[18] These population ratios by school and district can be aggregated to the state level, weighting by grade 1 enrollment.

Table 1 shows how these measures correlate at the state-year level. Not surprisingly, the CCD state population shares (1) and population ratios at the levels of state (2), district aggregated to state (3), and school aggregated to state (4) all correlate very highly, with $r > 0.95$. But each of the CCD

[17] The Surveillance, Epidemiology, and End Results (SEER) Program, http://seer.cancer.gov/, processes population data from the U.S. Census Bureau to be used in calculating rates of cancer incidence in the population at the state and county levels. It produces a more consistent population series over time than the Census estimates.
[18] Grissmer, Flanagan, Kawat, and Williamson (2000) employ this technique at the state level. At smaller geographies, there is a chance that this ratio exceeds unity, but empirically this occurred only in about 3 percent of cases. Functionally, we recoded ratios above 1 but less than 1.5 to unity, and we dropped observations with ratios of 1.5 or greater, although the results are not sensitive to these restrictions.

measures in turn also correlates highly with the NIEER state-funded pre-K rate, with $r > 0.75$. The CCD measures also correlate fairly strongly with the ACS public enrollment rate of 4-year-olds, with $r > 0.55$. Reassuringly, the CCD measures do *not* significantly correlate with NIEER's enrollment statistics for Head Start, most of which takes place outside public schools.[19] The CCD pre-K enrollments thus appear to have ample external validity.

Analytic Samples

Using the data sources described, we construct our analytic samples by merging the pre-K enrollment measures from CCD with NAEP data. Because students taking the 4th grade (8th grade) NAEP would have been enrolled in pre-K five (nine) school years earlier, assuming normal grade progression, our matching procedure incorporates this lag. Given the NAEP administrations for each state and subject and the availability of pre-K enrollment from CCD, Appendix Tables 2 and 3 show valid state-year combinations that compose the analytic samples.[20]

Because the NAEP data are at the student level and the CCD pre-K data—which provide the source of identifying variation—are at the school, district, or state level, we collapse the NAEP data to cells defined by NAEP test year, grade, test subject (math or reading), and geographic unit. We describe the details of this step in the empirical strategy section, below.[21] This produces samples at the state-year level, the district-year level, and the school-year level. While the NAEP data can be matched to CCD data for all test years at the state level, the matching at substate levels relies on the district and school identifiers in the restricted NAEP, which are missing in a few instances.[22] On average, our NAEP estimates for a given state, test year, grade, and subject in any of the samples is based on approximately 70 districts, about 140 schools, and about 3,100 students.

B. Methodology

[19] The correlation between the Census/ACS measure and NIEER's Head Start statistic is much higher, which is also plausible, as many families filling out the Census/ACS may consider Head Start as public school enrollment.

[20] A few states (and their constituent districts and schools) do not report pre-K enrollment in some years, which is the source of the blanks from 2003 onward. Notably, California never reports pre-K enrollment at the district or school level, or by race at the state level.

[21] Other researchers, notably Fitzpatrick (2008), estimated the effect of pre-K directly using the NAEP microdata. The advantage of this approach is the ability to control for student-level covariates. The two-step approach we employ retains much of this advantage while speeding up estimation.

[22] We successfully matched 100 percent of districts identified in the NAEP to the CCD, but because some schools in the NAEP lacked the school identifiers used in the CCD, we could match only 94 percent of NAEP schools (across all years) to the CCD.

Our augmented differences-in-differences strategy employs a two-stage design to estimate the effects of pre-K access on students' academic and behavioral outcomes. The first-stage uses the NAEP microdata to regress student-level outcomes on student-level covariates and a vector of geography-year indicator variables. The coefficients on these dummies, which represent means of the outcome variable adjusted for student characteristics, become the outcome variables for the second stage. The second stage, in turn, regresses these adjusted means on the appropriate pre-K measure and other covariates to identify the causal impact of pre-K. Donald and Lang (2007) demonstrate that such a two-stage approach can yield better inference when the number of groups is small; it also is computationally simpler.

More specifically we first estimate the equation:

$$y_{ig} = \boldsymbol{X_{ig}\alpha} + \boldsymbol{Z_g\gamma} + \varepsilon_{ig}\ , \tag{1}$$

where y_{ig} is a student-level test score, indicator variable for whether the student receives special education services, or indicator for above modal age for grade, with *i* indexing students and *g* indexing geography (state, district, or school). $\boldsymbol{X_{ig}}$ is a vector of student characteristics including binary indicators for sex, race, participation in the federal free lunch program, and participation in the federal reduced-price lunch program.[23] $\boldsymbol{Z_g}$ is a vector of indicator variables for geography. Finally, ε_{ig} is a student-level error term. Equation (1) is estimated separately for each NAEP year, subject (math or reading), and grade (4th or 8th), allowing the relationship between student characteristics and outcomes to vary over time and across subjects and grades.

The coefficient estimates $\boldsymbol{\gamma}$**,** which we stack across years for each subject and grade, are geography-specific fixed effects, net of student characteristics. We reparameterize this vector (within year and subject) by subtracting the overall weighted mean outcome for the entire sample so that the new vector represents deviations from the national mean (and thus sums to zero).

The reparameterized vector $\tilde{\boldsymbol{\gamma}}$ becomes the outcome variable in the second stage:

$$\gamma_{gt} = \beta_0 + \beta_1 \cdot PreK_{g,t-\ell} + \boldsymbol{\mu_g} + \boldsymbol{\varphi_t} + \boldsymbol{Q_{gt}\theta} + \nu_{gt}\ , \tag{2}$$

where $PreK_{g,t-\ell}$ is the measure of pre-K in geography g, lagged the appropriate number of years to correspond to when the test cohort would have been enrolled in pre-K, $\boldsymbol{\mu_g}$ is a vector of geography dummies, $\boldsymbol{\varphi_t}$ is a vector of test year dummies, $\boldsymbol{Q_{gt}}$ is a vector of time-varying characteristics of the geography, and ν_{gt} is an idiosyncratic error term, which we allow to be arbitrarily correlated within

[23] We plan to add additional student covariates, including additional controls for accommodations for students with disabilities or English language learners, in a later draft.

geography. Equation (2) is estimated separately by grade and subject. The coefficient of interest is β_1, which shows how the normalized outcome changes when the pre-K measure varies from 0 (no pre-K) to 1 (presumed to be full, or universal, pre-K).

We implement this two-step procedure at the level of states, districts, and schools. At the state level, we conduct three sets of analysis: one by race, one by the income classification of the schools, and one by the income classification of the students.[24] These levels of analysis are akin to the studies by Cascio and Schanzenbach (2013) and Rosinsky (2014) , except we use the CCD to measure pre-K intensity. Our major innovation, however, is to also examine the relationship between student outcomes and pre-K access at the district and school levels. In addition to providing significantly greater variation in pre-K than is possible with a state-level design (see descriptive statistics in Table 2), analysis at the district and school levels also allows greater control of possible unobservables that can bias estimates. As discussed in the first section, pre-K expansion at the state level may be correlated with the adoption of other state policies designed to boost achievement, especially since educational standards are set at the state level. Furthermore, even with a rich set of time-varying controls, it is plausible that unobserved variation remains. By exploiting within-state variation in pre-K, we can control not only for state and year fixed effects but their interaction, capturing the impacts of possibly nonlinear state-level policies that may be endogenous to pre-K expansion. Additionally, because many districts and schools are sampled multiple times across NAEP test years, we can include district (or school) fixed effects to net out permanent differences across these geographies.[25] Moreover, the CCD also permit us to control for time-varying characteristics of districts (or schools), including the share of students eligible for the assisted lunch program (categorical), racial and ethnic composition (categorical), and whether the district (school) is in an urban area, suburban area, town, or rural area.[26]

[24] For the race analysis, we use race-specific pre-K enrollment, as this is reliably available at the state level. For income classification by school, we use the CCD to determine the average share of students eligible for free or reduced-price lunch at each school over the entire sample period; schools with a share at or above 50 percent are classified as low-income, and schools with shares below 50 percent are classified as not low-income. We then aggregate these by state for matched schools in the NAEP. For income classification by student, we use the NAEP variable for free or reduced-price lunch to classify each tested student and aggregate by state.

[25] Due to the large number of districts and especially schools, we use the -reghdfe- package in Stata to implement the fixed effects.

[26] At the school-level, we also control for the teacher-student ratio. Although most schools and districts do not change urbanicity over time, some do due to development. All of these time-varying school and district controls are measured as of the NAEP test year, rather than an average over the years students would have been enrolled between pre-K and the NAEP test year, as in some previous studies. We not expect our results to be sensitive to this choice, as the results are not sensitive to including or excluding the time-varying controls.

For the state-level regressions, our identifying variation comes from changes in pre-K enrollment within state over time. For the district (and school) regressions, the identifying variation comes from within-district (within-school) changes in pre-K enrollment over time, relative to other districts (schools) in the same state. In the latter cases, in addition to examining average effects of pre-K across districts or schools, we also examine heterogeneity over certain types of districts or schools, specifically by the share of students on free or reduced-price lunch, by racial composition, by total enrollment, and by urban setting.[27]

Although we consider our approach to have several advantages over earlier studies, it is not without a set of disadvantages. First, even though we can better account for possible endogeneity in the pre-K expansion, we cannot eliminate it entirely. If individual districts expand pre-K because test scores are trending downward, our methodological approach will not adequately control for it.[28] We do not think this is likely to be a major problem, because in order for this to cause bias, the same time trends that caused districts to expand pre-K would have to be persistent enough to cause test score effects in the NAEP test year. It should be recalled that we always control for district or school fixed effects, thus avoiding endogeneity biases due to persistent levels of school or district characteristics being correlated with pre-K enrollment rates. In addition, we can control for state-year fixed effects, which avoids endogeneity bias due to state policy changes over time being correlated with state test score trends. Second, we do not account for pre-K program quality, including length of school day, as there is no measure of quality available for every district and school. Instead, our results will capture an average treatment effect of all public pre-K programs as they were actually implemented, and such an average treatment effect (even within certain district, school, or student types) may mask strong positive impacts from some programs and negative impacts from others. As a supplement to the main analysis, we do use NIEER quality metrics at the state-level to explore the possibility of heterogeneous pre-K impacts by quality, but this does not exploit intrastate variation.[29] Third, we do not capture the counterfactual, or other early childhood education programs for which public pre-K may substitute

[27] In each case, we use categorical indicators based on sample averages of the characteristic. These factors are among those that have been identified in previous research as showing heterogeneous treatment effects (Cascio and Schanzenbach 2013, Fitzpatrick 2008).

[28] Since NAEP results are not released by district, this is problematic only to the extent that NAEP results correlate with other state and district exams.

[29] We use the NIEER indicators, averaged over the sample horizon, to classify states into quality quartiles. In addition to these quartiles based on the simple sum index, we employ principal components analysis using the Stata package -polychoricpca- to account for correlation across the ten binary indicators. Finally, we create a third quality measure based on average real per-pupil spending, also provided by NIEER. As expected, the first two measures are highly correlated with each other (Spearman's $\rho = 0.81$), but the third is only weakly correlated with either of the first two (Spearman's $\rho = 0.42$ to 0.52).

(Feller et al. 2014; Kline and Walters 2015). Pre-K in the public schools may draw some students away from private programs, and others from governmental programs such as Head Start, and the availability of these programs, especially at substate levels, is hard to measure.[30]

IV. Results

A. *State-Level Results by Race*

Table 3 presents results from estimating equation (2) at the state-level, separately by race. Each of the four panels corresponds to a different test grade (4 or 8) and subject (math or reading. We examine four outcomes across columns: the (first-stage) adjusted scale score, the adjusted percentile score, the adjusted share of students reporting an Individual Educational Program (i.e., receiving special education services), and the adjusted share of students above modal age for grade. The rows designate different regressions for each race or ethnicity.

Looking broadly across all the estimates, we generally find small effects that are not statistically different from zero, and few systematic patterns.[31] Although white students in grade 4 show marginally significant *negative* effects of pre-K exposure on reading scores of modest size (5 scale score points or 4 percentiles), and slightly weaker negative effects on grade 8 reading, the coefficient estimates are essentially zero for math. Black students, on the other hand, consistently show beneficial (right-signed) effects of pre-K exposure across grades, subjects, and outcomes, although in only one case is the estimate statistically significant. Hispanics show no clear pattern, and students of other ethnicities also exhibit small point estimates that vary around zero.[32] In terms of magnitude, the point estimates are smaller than have been found in previous studies using state-level variation. On the one hand, we might expect larger point estimates, as many previous studies used a dichotomous indicator for pre-K while we use a continuous one, and even in states such as Georgia and Oklahoma that adopted large-scale public pre-K programs, participation among 4-year-olds was far from universal. Thus, as a matter of scaling alone, the estimates we show should approximately be halved to be commensurate with those from many of the earlier studies. On the other hand, it is quite possible that pre-K exposure averaged across

[30] It may be possible to capture private pre-K enrollment through the Department of Education's Private School Universe dataset, the analogue to the CCD for private schools. Head Start enrollment is available from historical Program Information Reports. We will pursue using these sources to control for the program choice set in future drafts.

[31] Allowing for state-specific linear time trends does not alter this conclusion. See Appendix Table 4.

[32] Although several estimates in grade 8 reading are statistically significant, we do not much stock into them, as it seems unlikely that no effect would be found at grade 4. Rather, these estimates may be a data artifact due to most states having relatively few "other" ethnicity students at grade 8 for most test years (note standard errors are much lower than for other races). They are also not robust to the inclusion of state time trends.

different quality programs yields smaller net effects. We note, however, that our estimates are of comparable precision to those in Fitzpatrick (2008) and Cascio and Schanzenbach (2013) once the Conley-Taber adjustments are applied. Put differently, we cannot rule out effect sizes of 5 to 8 percentiles for blacks, which is in the same ballpark as the earlier studies. Because such effect sizes can have large impacts on later outcomes, as shown earlier, there appears to be a limit on the usefulness of state-level variation in precisely identifying the impact of pre-K.

B. *State-Level Results by School and Student Income Proxies*

Nonetheless, we also report state-level results by school income (as proxied by share of students eligible for the assisted lunch program) and student income (as proxied by student eligibility in the lunch program). We do this to be consistent with previous studies, which have also used such breakdowns.

Table 4 shows results for both low-income and non-low-income schools and students. (Unlike the estimates in Table 3, which were based on race-specific pre-K enrollment shares of 4-year-olds, the estimates from here on are based on ratios of pre-K enrollment to grade 1 enrollment.) As with Table 3, there are few systematic patterns, and only a single estimate is statistically significant, and even then marginally so. That said, the point estimates are positive for math scores, consistent with previous studies, and magnitudes are larger for low-income students, also consistent with previous research. Although we cannot reject that pre-K has no effect on grade 4 math scores, we also cannot reject that it boosts them by 7 percentiles. The coefficient estimates tend to be negative for reading, and those for special education or overage bounce around zero. As before, the state-level variation is insufficient to rule out meaningfully-sized effects, positive or negative.

C. *District-Level Results*

We thus turn to estimates using district-level variation in pre-K. Since approximately 70 districts are sampled from each state on average, many of which are resampled for a future NAEP exam, the effective number of observations and identifying variation is much larger than in the state-level results.[33] These results are shown in Table 5. The first row of each panel shows the overall effect of pre-K, averaged across all school districts, and the estimates control for state-year dummies as well as

[33] Technically, schools are sampled, not districts. Thus, in several cases different schools within the same district are sampled over time. To the extent that there is significant variation in schools within a district (as is more likely with larger districts), estimation results may be confounded by compositional change. We attempt to address this issue by also controlling for school characteristics, in addition to individual characteristics, in the first stage. Furthermore, school-level results obviate this issue, but at the cost of not being able to estimate 8th grade outcomes, as very few schools have both pre-K and 8th grade.

district fixed effects and time-varying characteristics. Notably, the additional variation in pre-K yields much more precise estimates than those in Tables 3 or 4: standard errors are one-fifth to one-fourth the size. However, even these more powerful estimates generally do not indicate much of an effect of pre-K on test scores. For grade 4, the estimates show insignificant effects of -0.1 percentiles for math and -0.4 percentiles for reading. Recall again that these are estimated effects when a school district switches from having zero students in pre-K to having 100 percent of students enrolled in pre-K. Most observers would describe such effects as small. Note that these effects are even smaller when one considers more common variations in the scale of public pre-K (for example, the interquartile range in pre-K variation across districts is about 35 percentage points). Furthermore, the estimates are precise enough to rule out positive effects as small as 1 percentile, and negative effects of 2 percentiles, when moving from no public pre-K to universal public pre-K. Test score effects at grade 8, which rely on slightly different pre-K variation given the necessary lag time, show a qualitatively similar pattern.

Although there appear to be no meaningful average impact of pre-K on test scores, or on the likelihood of being overage for grade (column 4), there is a statistically significant and relatively large reduction in the propensity to receive special education services in grade 4. What's more, this result is consistent for both the math and reading NAEP subjects: the former shows a 1.3 percentage point decline and the latter a 0.9 percentage point decline, both relative to an (unadjusted) mean of 14.3 percent. The effect thus represents a 6–9 percent decline. Because the math and reading NAEP exams are administered to independently sampled schools and districts, this result appears quite robust. However, it does appear to fade out by 8th grade. The point estimates are still negative, but the magnitude is much smaller and not close to statistical significance.

Although this special education effect is worth considering for policy purposes, the magnitude of this effect is much smaller than in some other studies. For example, the Chicago Child-Parent Center is estimated to have reduced students who ever received special education services from 25 percent to 14 percent (Reynolds, Temple, White et al. 2011). Perry reduced total years in special education through high school from an average of 5.2 years to 4.0 years. Bagnato et al. (2009) found that Pennsylvania's pre-K program reduced special education services from 18 percent to 2.4 percent. The effects estimated in the current study are only a small fraction of these prior estimates.

The remaining rows of Table 5 represent estimation results for separate types of districts. As mentioned previously, an average result close to zero may mask heterogeneity across different programs, or in this case district characteristics. We classify districts by enrollment ethnicity, the share of students eligible for free or reduced-price lunch, district enrollment size, and district locale. These subgroup estimates are unsurprisingly noisier than the overall district-level estimates, but they are still more precise than the state-level subgroup results. When comparing so many estimates, some are bound to show statistical significance by chance, and so systematic patterns become more convincing. In particular, pre-K enrollment in smaller districts (fewer than 2500 students, about the national median) yields a negative impact on both math and reading grade 4 test scores, of about 1.5 percentiles, with confidence intervals bounding the effect approximately between 0 and −3 percentiles. The decline in special education services, on the other hand, appears to be concentrated in larger districts, with the estimate being statistically significant and virtually identical on both the math and reading tests. It is harder to be confident about the decline in special education services among poorer districts, as the statistically significant decline of 2.1 percentage points among districts participating in the math exam is not matched among districts taking the reading exam. A similar comparison casts doubt on the over-age results for majority-black and town districts. Grade 8 results are less clear, but the confidence intervals generally rule out a positive impact of pre-K on test scores.

To summarize, public pre-K expansion does appear to significantly reduce special education services at grade 4 on average, and particularly among larger districts. However, it does not appear to increase standardized test scores in math or reading, even for low-income districts. Indeed, we can rule out effects as small as 1 percentile when moving from no pre-K to universal pre-K, or less than 0.5 percentile when scaling to adjust for more typical pre-K expansions. To be sure, we cannot infer that no pre-K programs in our data produced positive test score impacts; what we can infer is that as implemented on average across districts and the types of districts investigated here, pre-K did not appreciably boost test scores. Nonetheless, these results speak to the importance of examining socioemotional and behavioral outcomes when evaluating pre-K, especially given the well-documented fadeout of effects on test scores.

D. *School-Level Results*

Examining the effects of pre-K at the school level offers the greatest degree of variation and at first blush the strongest opportunity to control for other factors that may influence outcomes. Whereas the district-level regressions may involve different sampled schools over the sample horizon, for which school-level controls can partially address composition bias, the school-level regressions can fully control for time-invariant school (or neighborhood) factors that may be related both to pre-K access and student outcomes. On the other hand, since few schools include both pre-K and 8th grade, it is not feasible to use a school-level analysis to investigate 8th grade outcomes. Also, perhaps more substantively, it is not always the case that students attend the same school for pre-K as they do for 4th grade. Some students and their families move across school boundaries in this interval, and some districts may consolidate their pre-K programs in just one or two elementary schools, especially if they are not universal and only a subset of students are eligible. Both issues would introduce measurement error and attenuation bias to the estimates. We thus view the analyses at the district and school levels as complements, with both having the potential to be informative.

Table 6 thus shows the school-level results, analogous to the district-level results in Table 5, except only for 4th grade. Consistent with the even greater variation across schools, standard errors are smaller than in Table 5. As with those earlier results, there appears to be no average effect of pre-K diffusion on math or reading test scores. The point estimates of 0.2 (math) and -0.1 (reading) percentiles not only are very small in magnitude, the confidence intervals are tight enough that effects of 1.1 percentiles (or 1 scale score point) can be ruled out. Similarly, the effects of pre-K on the likelihood of being overage for grade are minimal, and impacts of 1.5 percentage points in either direction are at or beyond the limits of the confidence intervals. Unlike the district-level results, the school-level analyses do not show a significant impact of pre-K on the probability of receiving special education services.[34] Although it is possible the discrepancy is due to a school composition issue, it could also arise from school-level heterogeneity, as the sample of matched schools does not nest exactly in the set of matched districts.

We again check for the possibility of heterogeneous effects, this time across school rather than district types. Although a few estimates are marginally statistically significant, it is likely that this is due to chance (type I error), as there are not consistent patterns in the estimates, and they are of relatively small magnitude. More generally, we do not find that the observed diffusion of pre-K meaningfully affected test scores, the likelihood of receiving special education services, or the likelihood of being overage at 4th grade for any of the school types classified by race, poverty, size, or locale. Even for these subsets of schools, our estimates are precise enough to bound maximum effects, in either direction, at 2–3 percentiles for test scores and 3–4 percentage points for special education and overage. In many cases, bounds are even smaller, especially when considering that the estimates represent the effect of moving from zero pre-K to pre-K that matches grade 1 enrollment, and the mean change across schools is only 19 percentage points (32 percentage points among schools with non-zero pre-K at some point).

The school-level results, despite a few differences, are broadly similar to the analyses that use pre-K variation at the district and state levels . Pre-K, as it has been implemented in the public schools

[34] The differences in the point estimates are statistically significant at the 5 percent level.

throughout the entire country has not meaningfully affected test scores, special education receipt, or grade retention. This inference holds on average across all students as well as for students of different races and socioeconomic levels, and across different types of districts and schools. While the state-level results are relatively imprecise, leaving open the possibility of effects as large as 10 percentiles (percentage points), the district- and school-level estimates shrink this window to nearly one-tenth the size. Although *certain* pre-K programs may have large effects on exposed student's academic and educational outcomes, the evidence here suggests that *typical* pre-K programs in the public schools have had essentially no effect on these outcomes by the end of elementary school.

E. *NIEER "Quality" Results (state-level)*

Finally, we also present results that attempt to control for the quality of the pre-K program. More specifically, Table 7 show results at the state level—the level at which NIEER produces quality benchmarks—for which the CCD pre-K measure has been interacted with quartiles of the NIEER quality index.[35] Assuming that program quality matters *and* that the NIEER measures adequately capture quality, it is possible that a null finding on average masks positive effects for higher-quality programs and perhaps a negative impact of lower-quality programs. The estimates in Table 7, however, do not provide evidence that this is the case. Although they are relatively noisy, the estimates if anything suggest the opposite gradient, with beneficial effects more likely in lower-quality quartiles. This pattern could be consistent with endogenous adoption of stronger pre-K programs in states that were negatively trending.[36] Alternatively, and as discussed previously, the NIEER benchmarks may be insufficient to measure quality. A more direct measure may be per-pupil spending, which NIEER also tracks. Appendix Table 6 presents similar results by quality quartiles, but using spending as a measure of quality. Although they are hardly precise, these results also do not show much of a quality gradient. They do, however, indicate beneficial effects of public pre-K expansion for black and Hispanic students—especially for mathematics—although the level of spending does not appear to matter much.[37] Test scores in math rise by a statistically significant 5 to 10 percentiles, a meaningfully large effect, although confidence intervals are rather wide, and effects as large as 15–20 percentiles, or as small as 2 percentiles, cannot be ruled out.

[35] See footnote 29 for how we construct the quartiles. We are exploring using substate variation and applying the state-level quality measures to district or school pre-K measures.

[36] Adding state-specific time trends produces qualitatively similar results, although the short panel (four observations per state) makes inference difficult.

[37] The point estimates across quality levels in Appendix Table 6 are larger than those in Table 3; the difference lies in the time period of the sample. Indirectly, pre-K program expansion since 2000 may have been higher quality than expansion during the 1990s. We plan to explore this possibility in future research.

V. Discussion and Conclusions

In this paper, we have used several datasets that together allow us to investigate the relationship between pre-K diffusion and educational outcomes on a representative sample of public schools and school districts throughout the country. Unlike most prior research, we do not examine the effects from a particular pre-K program or even a particular state's pre-K program; rather, we estimate the effects of all public pre-K programs averaged together. The approach we use has advantages over previous geographic studies of providing far more identifying variation, controlling for more covariates that were potentially unobserved confounders, and producing national-level estimates. This last advantage also extends to randomized control trials of pre-K, which typically yield concerns of external validity over whether they generalize to other settings and time periods. On the other hand, our approach also has disadvantages relative to earlier studies. We do not directly observe individual-level treatment or short-term outcomes, as in control trial studies. And relative to both the control trial studies and the geographic studies, our measure of treatment is diffused because we pool many different programs together and thus cannot separately identify the effect of a "high-quality" program from a mediocre or poor one. Put differently, whereas many prior studies looking at intensive or widely-regarded programs analyzed what a pre-K program *could do* under the right circumstances, in this paper we effectively look at what typical pre-K programs *have done* in practice over the last two decades.

Our results consistently indicate that pre-K programs in the public schools have done quite little, on average, when it comes to test scores, receiving special education services, or being retained in grade. The precision of results using district and school-level variation in pre-K generally rules out test score effects of about 1 percentile when moving from no pre-K to near-complete pre-K. As noted previously, a 1 percentile increase in test scores, all else equal, would be predicted, based on Chetty et al. (2011), to increase the present value of future earnings per individual by about $3,700 (for a 4th grade test score increase) to $5,800 (for an 8th grade test score increase). Quality pre-K is thought to cost at least $5,250 per student for a half-day school year program, and by at least $9,450 for a full-day school year program (Gault et al. 2008). The median of actual public pre-K spending since 2000, over both half-day and full-day programs, is about $4,400 (Barnett et al. 2014).[38] Therefore, our estimates do not imply that expected earnings benefits clearly exceed typical pre-K costs, let alone costs for "quality" programs. (Since we do not estimate the impact of "quality" programs, those costs are perhaps less relevant.)

[38] These figures are all in year 2011–2012 dollars.

We do find limited evidence (in the district-level results) that the expansion of pre-K in public schools reduced the assignment to special education services, by about 1 percentage point. These effects, if true, even at the upper bound of the confidence interval are not large enough alone to justify the costs of pre-K.[39] In calculating the special education cost implications of our estimates, it is reasonable to assume a year of special education might cost around $10,000.[40] If moving from zero to 100 percent of students in pre-K ended up reducing special education assignment by about 1 percentage point, special education costs per additional pre-K student would be lowered by about $100 per year of special education assignment. Even if this special education effect persisted for all 13 years from kindergarten through 12th grade, the cost savings per additional pre-K student, undiscounted, would add up to $1,300. As noted above, however, our estimates do not find effects on special education assignments at 8th grade, so there is little support for persistent effects on special education assignment for all 13 years. At best the savings on special education from expanded pre-K might save a few hundred dollars per student. In conjunction with the possible effect on test scores, these benefits of pre-K come close, but still do not exceed the costs in practice.

Thus, we believe a fair inference would be that the average public pre-K program in the United States may not be cost-effective, as measured by its impact on medium-term outcomes. We wish to caution readers that our findings do **NOT** imply that specific pre-K programs cannot be highly effective at boosting social outcomes over any given time horizon. Likewise, the findings also do **NOT** imply that public pre-K, on average, has no effect on long-term social outcomes.

First, we wish to strongly emphasize that we do *not* view the mostly null results found here as being in contradiction with the positive impacts found in several earlier studies. As noted, many of the previous pre-K studies concentrated on specific pre-K programs that were likely of higher-than-typical quality, as suggested both by expert opinion and by the magnitude of expenditures. Our results are instead broadly consistent with Rosinsky (2014), the only other study to our knowledge that looks at pre-K programs throughout the entire country.

---

[39] Some policymakers have been interested in using a possible lowering of special education costs as a way to finance pre-K. For example, a recent demonstration project in Utah involved “social impact bonds” in which Goldman Sachs paid for pre-K costs up front, and then was reimbursed by the government for the calculated reduction in special education assignments. However, the Utah project’s calculations of the magnitude of this special education cost reduction have been challenged by early education scholars (Popper 2015).

[40] Current expenditures per student in K–12 average about $11,000 per student in the 2011–12 year, calculated in year 2013–14 dollars (U.S. Department of Education, 2015b). Special education costs are estimated to be about 90 percent greater than regular education costs (Aron and Lopreset 2012). Even if none of the $11,000 per student reflected special education costs, the extra costs due to special education would be about $9,900. Since the $11,000 figure includes some special education, however, actual costs of regular education per student are somewhat lower, and the marginal cost of special education per student is thus somewhat less than $9,900.

Second, even if the average pre-K program produces no measureable impact on 4^{th} grade test scores, whether due to varied quality, test score fadeout, or both, it does not necessarily follow that there are no long-term, or “sleeper” effects. As Heckman has noted on multiple occasions, pre-K may boost long-term social outcomes as much (if not more) through its effect on socioemotional skills as on academic ones. If these soft skills are not adequately captured in our NAEP proxies of special education and overage for grade, future educational attainment and future earnings might be more greatly affected than predicted based on the medium-term results in the current paper.

We thus do not consider the current draft a complete one. Rather we consider several next steps. First, we plan more to employ more complete student-level controls and add additional district-level covariates, including expenditures and measures of pre-K enrollment in private schools and Head Start. This would act both as an additional robustness check and possibly shed light on the influence of alternative early childhood arrangements in mitigating pre-K impacts. Second, we plan to exploit the district and school variation *within* states considered to have effective programs, relative both to themselves and other states. This step would both generalize the identification strategy of Ladd, Muschkin, and Dodge (2014) to multiple states and allow for a more thorough investigation into program quality. Third, it is possible that the lack of average effects, even across district and school types, may mask effects from the weakest students. It should be possible to leverage the NAEP microdata to estimate quantile treatment effects, or how pre-K affects the test score distribution and not just its conditional mean. Finally, we plan to investigate the possibility of sleeper effects more directly by examining high school graduation outcomes at the district level using publicly available data from the Department of Education.

References

Aron, Laudan, and Pamela Loprest. 2012. "Disability and the Education System." *The Future of Children* 22(1): 97–122.

Bagnato, Stephen J. , Jennifer Salawy, and Hoi Suen. 2009. *Pre-K Counts in Pennsylvania for Youngster's Early School Success.* 2009 Final Research Report. Pittsburgh, PA: Early Childhood Partnerships.

Bailey, Martha J., and Susan M. Dynarski. 2011. "Inequality in Postsecondary Education." In G. J. Duncan and R. J. Murnane (eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage. pp. 117–132.

Barnett, W. Steven, Megan E. Carolan, Jen Fitzgerald, and James H. Squires. 2014. *The State of Preschool 2014: State Preschool Yearbook*. New Brunswick, NJ: Rutgers, Graduate School of Education, National Institute for Early Education Research.

Barnett, W. Steven, and Jason T. Hustedt. 2011. *Improving Public Financing for Early Learning Programs.* NIEER Policy Brief 23.

Bartik, Timothy J. 2011. *Investing in Kids: Early Childhood Programs and Local Economic Development*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bartik, Timothy J. 2014. *From Preschool to Prosperity: The Economic Payoff to Early Childhood Education*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bartik, Timothy J., William T. Gormley, Jr., and Shirley Adelstein. 2012. "Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups." *Economics of Education Review* 31(6): 1143–1161.

Belfield, Clive R., Milagros Nores, Steve Barnett, and Lawrence Schweinhart. 2006. "The High/Scope Perry Preschool Program: Cost–Benefit Analysis Using Data from the Age-40 Followup." *Journal of Human Resources*. 41(1): 162–190.

Burchinal, M., K. Kainz, K., and Y. Cai. 2011. "How Well Do Our Measures of Quality Predict Child Outcomes? A Meta-Analysis and Coordinated Analysis of Data from Largescale Studies of Early Childhood Settings. In M. Zaslow (ed.), *Reasons to Take Stock and Strengthen Our Measures of Quality*. Baltimore, MD: Brooks Publishing.

Camilli, Gregory, Sadako Vargas, Sharon Ryan, and W. Steven Barnett. 2010. "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teachers College Record* 112(3): 579–620.

Campbell, F. A., E. P. Pungello, M. Burchinal, K. Kainz, Y. Pan, B. H. Wasik, O. A. Barbarin, J. J. Sparling, and C. T. Ramey. 2012. "Adult Outcomes as a Function of an Early Childhood Educational Program: An Abecedarian Project Follow-up." *Developmental Psychology* 48(4): 1033–1043.

Carolan, Megan. 2013. "Problematic Pre-K Data in the U.S. Census, Part 1" and "Part 2". New America Foundation. Part 2 and link to Part 1 at: http://earlyed.newamerica.net/blogposts/2013/problematic_pre_k_data_in_the_us_census_part_2-77290

Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." NBER working paper 19735. Cambridge, MA: National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593–1660.

Chingos, Matthew M. 2015. *Breaking the Curve: Promises and Pitfalls of Using NAEP Data to Assess the State Role in Student Achievement*. Urban Institute Research Report.

Conley, Timothy, and Christopher Taber. 2011. "Inference with 'Difference-in-Differences' with a Small Number of Policy Changes". *Review of Economics and Statistics* 93(1): 113–125.

Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85(3): 341–364.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111–134. doi:10.1257/app.1.3.111

Donald, Stephen G., and Kevin Lang. 2007. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics*. 89(2): 221–233.

Duckworth, Angela L., and David S. Yeager. 2015. "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes." *Educational Researcher* 44(4): 237–251.

Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27(2): 109–132.

Feller, Avi, Todd Grindal, Luke Miratrix, Lindsay Page. 2014. "Compared to What? Variation in the Impact of Early Childhood Education by Alternative Care-Type Settings." Harvard University Working Paper.

Fitzpatrick, Maria D. 2008. "Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement." *B.E. Journal of Economic Analysis & Policy* 8(1) Advances, Article 46: 1–38.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4): 999–1012.

Gault, Barbara, Anne W. Mitchell, Erica Williams, Judy Dey, and Olga Sorokina. 2008. *Meaningful Investments in Pre-K: Estimating the Costs of Quality Programs*. Washington, D.C.: Institute for Women's Policy Research.

Grissmer, David, Ann Flanagan, Jennifer Kawata, and Stephanie Williamson. 2000. *Improving Student Achievement: What State NAEP Test Scores Tell Us*. Santa Monica, CA: RAND Corporation.

Guernsey, Lisa. 2013. "Comments on State of Preschool Survey 2013-2015". New America Foundation. Available at http://www.newamerica.net/publications/resources/2013/comments_on_state_of_preschool_survey_2013_2015.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rifkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18(5): 527–544.

Heckman, James J. 2015. "Quality Early Childhood Education: Enduring Benefits." October 15. Available at http://heckmanequation.org/content/quality-early-childhood-education-enduring-benefits. Accessed on October 28, 2015.

Heckman, James J., and Tim Kautz. 2012. "Hard Evidence on Soft Skills." *Labour Economics* 19(4): 451–464.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the High/Scope Perry Preschool Program." *Journal of Public Economics* 94(1-2): 114–128.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6): 2052–2086.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper No. 7867.

Kearney, Melissa S., and Phillip B. Levine. 2015. "Early Childhood Education by MOOC: Lessons from Sesame Street." NBER Working Paper No. 21229.

Keys, Tran D., George Farkas, Margaret R. Burchinal, Greg J. Duncan, Deborah L. Vandell, Weilin Li, Erik A. Ruzek, and Carollee Howes. 2013. "Preschool Center Quality and School Readiness: Quality

Effects and Variation by Demographic and Child Characteristics." *Child Development* 84(4): 1171–1190.

Kline, Patrick, and Christopher Walters. 2015. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." NBER Working Paper No. 21658.

Ladd, Helen F., Clara G. Muschkin, and Kenneth A. Dodge. 2014. "From Birth to School: Early Childhood Initiatives and Third-Grade Outcomes in North Carolina." *Journal of Policy Analysis and Management* 33(1): 162–187. doi:10.1002/pam.21734

Larsen, J. M. , & Robinson, C. C. 1989. "Later Effects of Preschool on Low-Risk Children." *Early Childhood Research Quarterly*(4): 133–144.

Lipsey, Mark W., Dale C. Farran, and Kerry G. Hofer. 2015. "A Randomized Control Trial of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third Grade." Vanderbilt University Working Paper.

Lipsey, Mark W., Christina Weiland, Hirokazu Yoshikawa, Sandra Jo Wilson, and Kerry G. Hofer. 2015. "The Prekindergarten Age-Cutoff Regression-Discontinuity Design Methodological Issues and Implications for Application." *Educational Evaluation and Policy Analysis* 37(3): 296–313.

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159–208.

Magnuson, Katherine A., Christopher Ruhm, and Jane Waldfogel. 2007. "Does Prekindergarten Improve School Preparation and Performance?" *Economics of Education Review* 26(1): 33–51.

Minervino, Jim. 2014. *Lessons from Research and the Classroom: Implementing High-Quality Pre-K that Makes a Difference for Young Children*. Seattle, WA: Bill and Melinda Gates Foundation.

Neidell, Matthew, and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *The Review of Economics and Statistics* 92(3): 562–576.

Olsen, Edgar, Paul Carillo, and Dirk Early. 2012. "A Panel of Price Indices for Housing, Other Goods, and All Goods for All Areas in the United States 1982-2008". University of Virginia working paper. Updated through 2012 at http://eoolsen.weebly.com/price-indices.html.

Oreopoulos, Philip, Marianne Page, and Ann Stevens. 2006. "Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling." *Journal of Labor Economics* 24(4): 729–760.

Popper, Nathaniel. 2015. "Success Metrics Questioned in Program Funded by Goldman." *New York Times*. November 3.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer. 2012. *Third Grade Follow-Up to the Head Start Impact Study: Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." In G. J. Duncan and R. J. Murnane (eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage. pp. 91–116.

Reynolds, Arthur J., Judy A. Temple, Barry A.B. White, Suh-Ruu Ou, and Dylan L. Robertson. 2011. "Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program." *Child Development* 82(1): 379–404.

Rosinsky, Kristina L. 2014. "The Relationship Between Publicly Funded Preschool and Fourth-Grade Math Test Scores: A State-Level Analysis."

Siegel, Robert. 2015. "New York City Mayor Goes All-In On Free Preschool." *NPR*. September 9.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: HighScope Educational Research Foundation.

Sabol, Terri J., S.L. Soliday Hong, R.C. Pianta, and M.R. Burchinal. 2013. “Can Rating Pre-K Programs Predict Children’s Learning?” *Science.* 341 (23): 845–846.

Smith, Alex. 2015. “The Long-Run Effects of Universal Pre-K on Criminal Activity.” U.S. Military Academy Working Paper.

U.S. Department of Education. 2015a. *National Assessment of Educational Progress (NAEP), various years, 1990–2015 Mathematics Assessments*. Washington, DC: Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education. 2015b. *The Condition of Education.* Washington, DC: U.S. Department of Education.

U.S. Department of Health and Human Services. 2010. *Head Start Impact Study* (Final Report). Washington, DC: DHSS, Administration for Children and Families.

Weiland, Christina, and Hirokazu Yoshikawa. 2013. “Impacts of a Prekindergarten Program on Children’s Mathematics, Language, Literacy, Executive Function, and Emotional Skills.” *Child Development* 84(6): 2112–2130.

Yoshikawa, Hirokazu, Christina Weiland, Jeanne Brooks-Gunn, Margaret R. Burchinal, Linda M. Espinosa, William T. Gormley, Jens Ludwig, Katherine A. Magnuson, Deborah Phillips, and Martha J. Zaslow. 2013. *Investing in Our Future: The Evidence Base on Preschool Education*. Available at http://fcd-us.org/resources/evidence-base-preschool or http://www.srcd.org/policy-media/policy-updates/meetings-briefings/investing-our-future-evidence-base-preschool.

Zaslow, Martha, Rachel Anderson, Zakia Redd, Julia Wessel, Louisa Tarullo, and Margaret Burchinal. 2010. *Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Table 1 Correlations of Pre-K Measures Across Data Sources, at State-Year Level

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	CCD State PK Share of 4-year-olds	**1.000**						
(2)	CCD State PK/G1 ratio	0.976	**1.000**					
(3)	CCD District PK/G1 ratios (aggregated)	0.957	0.981	**1.000**				
(4)	CCD School PK/G1 ratios (aggregated)	0.949	0.962	0.980	**1.000**			
(5)	NIEER State PK Share of 4-year-olds	0.768	0.752	0.759	0.821	**1.000**		
(6)	NIEER Head Start Share of 4-year-olds	0.098	0.060	0.052	0.090	0.195	**1.000**	
(7)	Census/ACS Share of 4-year-olds	0.559	0.582	0.590	0.592	0.600	0.396	**1.000**

SOURCES: Authors' calculations from the Common Core of Data (various years), NIEER State Preschool Yearbooks (various years), 1990 and 2000 Census and American Community Surveys (various years).

NOTE: Pairwise Pearson correlations are calculated at the state-year level for all valid state-year pairs. CCD data cover fall 1990 through fall 2007 school years, NIEER data cover fall 2001 through fall 2007 school years, and Census/ACS data cover spring 1990 (matched to fall 1990 in CCD), spring 2000 (matched to fall 1999 in CCD), and fall 2001 through fall 2007. The ACS enrollment share matched to the fall of each year *t* is a weighted average of the ACS fielded in year *t* (0.375) and year *t+1* (0.625) to approximate coverage for the school year. CCD ratios are calculated by summing the numerator within unit, summing the denominator within unit, taking the quotient, and then averaging using the denominator as weights. We do not use data beyond the fall of 2007, as that is the latest year that can be matched to 4^{th} grade outcomes in NAEP.

Table 2A Summary Statistics for State-Level Race Samples, Grade 4

	White		*Black*		*Hispanic*		*Other*	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre-K share	0.163	0.122	0.341	0.246	0.296	0.240	0.181	0.182
Math Scale Score, raw	245.4	6.14	219.9	7.88	227.7	7.91	237.2	14.10
Math Scale Score, student-adjusted	3.87	3.33	-11.86	5.40	-4.34	6.23	-0.20	10.14
Math Percentile Score, raw	54.4	6.41	29.11	6.33	36.51	6.88	46.39	13.39
Math Percentile Score, student-adjusted	3.81	3.47	-11.64	4.16	-4.57	5.44	0.08	9.53
Reading Scale Score, raw	227.7	5.21	200.9	7.72	205.3	7.70	214.7	15.66
Reading Scale Score, student-adjusted	4.35	3.61	-9.91	5.61	-6.03	6.72	-3.30	11.98
Reading Percentile Score, raw	54.90	4.48	32.75	5.35	36.68	5.45	44.97	12.00
Reading Percentile Score, student-adjusted	3.57	3.05	-8.42	3.64	-4.93	4.70	-2.06	8.93
Special Ed (math), raw	0.142	0.024	0.156	0.039	0.134	0.036	0.119	0.044
Special Ed (math), student-adjusted	0.203	0.023	0.004	0.041	-0.016	0.035	-0.014	0.038
Special Ed (reading), raw	0.140	0.024	0.155	0.039	0.137	0.034	0.120	0.042
Special Ed (reading), student-adjusted	0.018	0.023	0.005	0.040	-0.014	0.034	-0.014	0.036
Overage for Grade (math), raw	0.411	0.092	0.421	0.106	0.416	0.088	0.331	0.149
Overage for Grade (math), student-adjusted	0.039	0.088	0.014	0.102	0.008	0.086	-0.055	0.143
Overage for Grade (reading), raw	0.413	0.091	0.422	0.105	0.417	0.083	0.334	0.149
Overage for Grade (reading), student-adjusted	0.040	0.087	0.014	0.101	0.010	0.084	-0.055	0.143

NOTE: All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. Sample sizes in the second stage (at the state-year level for each race) are between 310 and 350, depending on subject. Cell sizes in the first stage—the number of students contributing to the outcome mean at the state-year level for each race—average about 1,910 for whites (min=50, max=3,160), 580 for blacks (min=0, max=2,820), 500 for Hispanics (min=0, max=5,930), and 260 for others (min=0, max=2,830). All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

Table 2B Summary Statistics for State-Level Income Samples, Grade 4

Variable	*Not Low-Income School*		*Low-Income School*		*Not Low-Income Student*		*Low-Income Student*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre-K share	0.162	0.130	0.268	0.216	0.187	0.146	0.220	0.176
Math Scale Score, raw	244.7	6.70	226.2	8.72	246.5	7.52	225.1	8.78
Math Scale Score, student-adjusted	2.30	3.86	-4.71	5.59	6.79	4.31	-9.43	5.33
Math Percentile Score, raw	53.51	6.77	35.47	7.18	55.43	7.73	34.31	7.23
Math Percentile Score, student-adjusted	2.14	3.90	-4.61	4.79	6.72	4.44	-9.38	4.44
Reading Scale Score, raw	228.3	4.82	206.4	6.95	230.9	5.27	204.7	7.03
Reading Scale Score, student-adjusted	3.02	4.54	-5.60	5.84	8.79	4.56	-11.47	5.67
Reading Percentile Score, raw	55.41	4.09	37.47	4.84	57.61	4.58	36.01	4.91
Reading Percentile Score, student-adjusted	2.44	3.81	-4.61	3.98	7.31	3.92	-9.50	3.82
Special Ed (math), raw	0.135	0.024	0.147	0.032	0.110	0.023	0.167	0.039
Special Ed (math), student-adjusted	0.015	0.023	-0.001	0.032	-0.027	0.021	0.038	0.036
Special Ed (reading), raw	0.136	0.023	0.155	0.039	0.110	0.023	0.168	0.039
Special Ed (reading), student-adjusted	0.015	0.022	0.005	0.040	-0.027	0.020	0.037	0.036
Overage for Grade (math), raw	0.387	0.098	0.433	0.097	0.371	0.097	0.433	0.107
Overage for Grade (math), student-adjusted	0.009	0.094	0.034	0.095	-0.022	0.095	0.049	0.104
Overage for Grade (reading), raw	0.385	0.099	0.431	0.097	0.369	0.098	0.434	0.107
Overage for Grade (reading), student-adjusted	0.010	0.094	0.036	0.095	-0.022	0.096	0.051	0.105

NOTE: Low-income schools are those where the share of students receiving free or reduced-price lunch exceed 50 percent on average over the sample horizon according to the CCD; non-low-income schools are the converse. Low-income students are those identified as such in the NAEP data. All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. Sample sizes in the second stage (at the state-year level for each group) are between 310 and 390, depending on subject. Cell sizes in the first stage—the number of students contributing to the outcome mean at the state-year level for each group—average about 1,810 for non-low-income schools (min=50, max=3,200), 1,420 for low-income schools (min=0, max=6,820), 1,640 for non-low-income students (min=40, max=3,990), and 1,580 for low-income students (min=20, max=7,060). All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

Table 2C Summary Statistics for District-Level Samples, Grade 4

Variable	*Pre-K Share*		*Scale Score*		*Percentile Score*		*Special Ed*		*Overage*		*2nd stage N*	*Unique districts*
Math	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
Overall	0.209	0.231	237.3	13.43	46.26	13.08	0.139	0.069	0.408	0.132	19,320	5,280
Majority-Black Districts	0.351	0.295	220.1	11.38	30.03	9.58	0.135	0.060	0.425	0.121	1,140	300
Majority-Hispanic Districts	0.393	0.288	230.5	10.20	39.29	9.26	0.125	0.053	0.441	0.096	800	200
90+% White Districts	0.153	0.219	244.1	10.88	52.89	11.43	0.145	0.090	0.420	0.153	7,970	2,270
40+% FRL Districts	0.268	0.261	230.7	12.00	39.71	11.14	0.140	0.068	0.429	0.128	9,300	2,520
<40% FRL Districts	0.139	0.164	245.1	10.47	54.02	10.76	0.137	0.069	0.384	0.132	10,000	2,740
2500+ Student Districts	0.216	0.224	236.8	13.34	45.79	12.93	0.138	0.060	0.398	0.122	10,180	2,480
<2500 Student Districts	0.176	0.259	239.7	13.59	48.47	13.56	0.143	0.100	0.445	0.160	9,140	2,800
City Districts	0.286	0.265	232.3	13.46	41.53	12.62	0.140	0.053	0.409	0.102	2,430	520
Suburb Districts	0.162	0.164	241.9	12.28	50.85	12.32	0.135	0.062	0.358	0.133	4,610	1,230
Town Districts	0.184	0.224	236.9	11.97	45.60	11.76	0.144	0.077	0.461	0.120	3,510	940
Rural Districts	0.173	0.237	238.3	13.22	47.08	13.12	0.138	0.094	0.449	0.149	8,770	2,590
Reading												
Overall	0.200	0.226	218.3	14.72	47.26	11.91	0.138	0.068	0.407	0.133	21,460	5,520
Majority-Black Districts	0.344	0.291	200.4	11.86	32.89	8.68	0.132	0.062	0.423	0.126	1,270	310
Majority-Hispanic Districts	0.371	0.284	208.0	11.24	38.65	8.33	0.125	0.054	0.442	0.094	850	200
90+% White Districts	0.147	0.214	227.0	11.50	54.21	9.96	0.143	0.087	0.420	0.150	8,880	2,420
40+% FRL Districts	0.257	0.255	210.4	12.70	40.73	9.73	0.139	0.067	0.425	0.132	10,080	2,560
<40% FRL Districts	0.133	0.164	227.4	11.20	54.76	9.53	0.137	0.069	0.386	0.132	11,330	2,940
2500+ Student Districts	0.205	0.219	217.7	14.50	46.74	11.75	0.137	0.060	0.396	0.123	11,530	2,580
<2500 Student Districts	0.172	0.257	221.5	15.32	49.77	12.36	0.140	0.096	0.446	0.158	9,920	2,930
City Districts	0.278	0.260	211.8	13.86	41.96	10.99	0.140	0.054	0.408	0.104	2,720	530
Suburb Districts	0.152	0.160	224.0	13.32	51.94	11.12	0.134	0.061	0.358	0.133	5,370	1,320
Town Districts	0.179	0.224	217.5	13.27	46.46	10.53	0.142	0.077	0.462	0.124	3,960	1,000
Rural Districts	0.167	0.235	219.8	14.97	48.33	11.96	0.137	0.091	0.449	0.148	9,400	2,670

NOTE: All outcome variables are unadjusted for student-level covariates. See text for details on construction of (CCD-defined) subgroups. All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. Average cell sizes in the first stage—the number of students contributing to the outcome mean at the district-year level for each group—range from about 20 for rural districts and 90+% white districts to about 120 for city districts; the overall average is about 40. All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

Table 2D Summary Statistics for School-Level Samples, Grade 4

Variable	*Pre-K Share*		*Scale Score*		*Percentile Score*		*Special Ed*		*Overage*		*2nd stage N*	*Unique schools*
Math	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
Overall	0.163	0.267	236.7	16.08	45.69	15.58	0.139	0.096	0.404	0.156	26,930	9,130
Majority-Black Schools	0.306	0.326	217.3	14.40	27.18	11.80	0.140	0.101	0.433	0.162	3,120	1,060
Majority-Hispanic Schools	0.287	0.338	225.9	12.59	34.81	11.30	0.132	0.090	0.430	0.136	2,500	880
90+% White Schools	0.125	0.241	244.4	11.46	53.17	12.04	0.142	0.104	0.419	0.160	8,420	2,660
50+% FRL Schools	0.237	0.311	225.6	13.90	34.77	12.35	0.145	0.101	0.430	0.157	12,110	4,220
<50% FRL Schools	0.103	0.207	245.5	11.64	54.39	12.02	0.135	0.092	0.383	0.153	14,780	4,900
400+ Student Schools	0.161	0.256	237.1	15.93	46.08	15.56	0.133	0.084	0.396	0.146	13,380	4,730
<400 Student Schools	0.167	0.287	235.9	16.35	44.89	15.59	0.151	0.117	0.420	0.174	13,550	4,400
City Schools	0.227	0.312	231.5	17.70	40.77	16.68	0.140	0.096	0.407	0.146	8,820	2,840
Suburb Schools	0.107	0.203	241.4	14.89	50.33	14.83	0.135	0.087	0.354	0.148	5,430	2,040
Town Schools	0.112	0.216	237.9	12.56	46.58	12.53	0.143	0.093	0.435	0.156	3,250	1,070
Rural Schools	0.153	0.257	238.9	14.19	47.66	14.00	0.141	0.108	0.436	0.167	9,430	3,180
Reading												
Overall	0.155	0.260	217.7	18.22	46.75	14.58	0.138	0.094	0.403	0.155	30,150	9,990
Majority-Black Schools	0.305	0.322	197.0	15.70	29.99	10.91	0.137	0.100	0.433	0.162	3,570	1,180
Majority-Hispanic Schools	0.275	0.330	202.8	15.08	34.66	11.02	0.131	0.089	0.432	0.137	2,770	940
90+% White Schools	0.119	0.234	227.3	12.43	54.51	10.76	0.140	0.099	0.419	0.159	9,260	2,880
50+% FRL Schools	0.225	0.304	204.6	15.58	36.07	11.32	0.144	0.099	0.432	0.157	13,690	4,630
<50% FRL Schools	0.100	0.204	228.1	12.65	55.25	10.83	0.134	0.088	0.380	0.150	16,410	5,350
400+ Student Schools	0.153	0.250	217.7	18.16	46.77	14.62	0.133	0.084	0.396	0.147	15,650	5,370
<400 Student Schools	0.160	0.281	217.7	18.34	46.72	14.50	0.151	0.111	0.420	0.172	14,510	4,620
City Schools	0.218	0.306	211.1	19.63	41.45	15.49	0.139	0.094	0.406	0.146	9,840	3,080
Suburb Schools	0.102	0.199	223.6	16.61	51.62	13.69	0.135	0.085	0.356	0.146	6,490	2,370
Town Schools	0.109	0.213	218.2	14.51	47.03	11.65	0.143	0.090	0.436	0.157	3,560	1,140
Rural Schools	0.147	0.252	220.4	16.41	48.87	12.99	0.138	0.103	0.436	0.165	10,260	3,400

NOTE: All outcome variables are unadjusted for student-level covariates. See text for details on construction of (CCD-defined) subgroups. All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. Average cell sizes in the first stage—the number of students contributing to the outcome mean at the school-year level for each group—range from about 10 for small schools and rural schools to about 30 for big schools and city schools; the overall average is about 20. All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

Table 3 The Effects of Pre-K by Race, Using State-Level Variation

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Whites	-0.97	-0.63	-0.002	0.007
	(1.71)	(1.79)	(0.013)	(0.032)
Blacks	7.39*	5.44	-0.004	-0.011
	(4.19)	(3.39)	(0.026)	(0.040)
Hispanics	3.35	2.23	-0.002	-0.005
	(4.18)	(2.87)	(0.029)	(0.055)
Others	3.38	2.85	0.016	-0.014
	(4.03)	(3.07)	(0.022)	(0.026)
Panel B: Grade 4 Reading				
Whites	-5.43*	-4.11*	0.008	0.028
	(2.74)	(2.19)	(0.012)	(0.035)
Blacks	2.15	1.45	-0.006	-0.061
	(4.08)	(2.82)	(0.024)	(0.047)
Hispanics	-3.08	-3.23	-0.018	-0.011
	(7.50)	(5.11)	(0.032)	(0.046)
Others	-0.09	-0.42	-0.005	-0.010
	(1.78)	(1.23)	(0.011)	(0.022)
Panel C: Grade 8 Math				
Whites	0.28	0.50	-0.005	0.020
	(2.83)	(2.26)	(0.020)	(0.038)
Blacks	5.34	3.00	0.012	-0.030
	(4.34)	(2.64)	(0.020)	(0.035)
Hispanics	-3.16	-1.68	-0.010	-0.051
	(4.55)	(3.37)	(0.038)	(0.060)
Others	-1.56	-1.65	0.025	-0.038
	(2.63)	(1.43)	(0.015)	(0.016)
Panel D: Grade 8 Reading				
Whites	-3.83	-2.90	-0.008	0.026
	(3.56)	(2.99)	(0.024)	(0.043)
Blacks	4.47	2.27	-0.018	0.003
	(3.01)	(2.04)	(0.019)	(0.040)
Hispanics	-4.49	-4.03	-0.021	-0.059
	(5.58)	(4.45)	(0.026)	(0.063)
Others	-4.84***	-3.53**	0.043*	-0.040**
	(1.63)	(1.52)	(0.022)	(0.013)

NOTE: Each cell is from a separate regression of the outcome on the (race-specific) pre-K measure, a set of state dummies, and a set of test year dummies. Each observation is a state-year, and there are 314 observations for each regression in panel A, 354 observations in panel B, 267 observations in panel C, and 306 observations in panel D. Standard errors in parentheses are clustered by state.

Table 4 The Effects of Pre-K by School and Student Income, Using State-Level Variation

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Low-Income Schools	2.84	1.86	0.006	0.010
	(3.34)	(2.50)	(0.027)	(0.030)
Non-low-income Schools	2.16	1.91	-0.002	0.072*
	(1.79)	(1.72)	(0.021)	(0.040)
Low-Income Students	4.46	3.23	0.003	-0.003
	(3.37)	(2.12)	(0.020)	(0.030)
Non-low-income Students	2.89	2.46	0.013	0.037
	(3.02)	(2.79)	(0.015)	(0.025)
Panel B: Grade 4 Reading				
Low-Income Schools	-1.95	-2.01	0.013	-0.015
	(3.62)	(2.45)	(0.021)	(0.030)
Non-low-income Schools	-3.83	-1.62	-0.003	0.014
	(3.17)	(2.77)	(0.012)	(0.049)
Low-Income Students	-2.37	-1.36	0.011	-0.027
	(3.90)	(2.49)	(0.018)	(0.030)
Non-low-income Students	-2.19	-2.08	0.010	0.004
	(3.84)	(3.18)	(0.013)	(0.032)
Panel C: Grade 8 Math				
Low-Income Schools	1.99	1.27	0.008	-0.040
	(4.57)	(2.91)	(0.025)	(0.028)
Non-low-income Schools	1.36	1.23	-0.012	0.048
	(4.01)	(3.11)	(0.026)	(0.052)
Low-Income Students	1.86	1.83	0.001	-0.024
	(3.37)	(2.21)	(0.024)	(0.024)
Non-low-income Students	-0.73	-0.24	0.013	0.008
	(3.44)	(2.58)	(0.016)	(0.041)
Panel D: Grade 8 Reading				
Low-Income Schools	-0.24	-1.59	0.006	-0.002
	(3.96)	(2.58)	(0.032)	(0.044)
Non-low-income Schools	-2.94	-1.98	-0.025	0.061
	(3.79)	(3.27)	(0.028)	(0.049)
Low-Income Students	0.29	-0.12	0.023	-0.003
	(3.53)	(2.42)	(0.028)	(0.045)
Non-low-income Students	-2.78	-2.35	-0.003	0.019
	(3.32)	(2.83)	(0.017)	(0.036)

NOTE: Each cell is from a separate regression of the outcome on the pre-K measure, a set of state dummies, and a set of test year dummies. Each observation is a state-year, and there are 353 observations for each regression in panel A, 359 observations in panel B, 308 observations in panel C, and 319 observations in panel D. Standard errors in parentheses are clustered by state.

Table 5 The Effects of Pre-K, Using District-Level Variation

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Overall	-0.18	-0.11	-0.013**	-0.001
	(0.54)	(0.53)	(0.006)	(0.007)
Majority-Black Districts	0.45	1.88	-0.023	-0.073**
	(2.43)	(2.13)	(0.022)	(0.034)
Majority-Hispanic Districts	-2.65	-2.78	-0.022	-0.020
	(2.63)	(2.54)	(0.025)	(0.039)
90+% White Districts	-0.93	-0.86	-0.011	0.004
	(0.88)	(0.91)	(0.009)	(0.012)
40+% FRL Districts	-0.22	-0.17	-0.021***	-0.009
	(0.72)	(0.69)	(0.007)	(0.010)
<40% FRL Districts	-0.14	-0.11	-0.003	0.006
	(0.83)	(0.87)	(0.008)	(0.011)
2500+ Student Districts	0.90	0.94	-0.014*	-0.006
	(0.78)	(0.75)	(0.008)	(0.011)
<2500 Student Districts	-1.69**	-1.64**	-0.013	0.005
	(0.71)	(0.73)	(0.008)	(0.011)
City Districts	1.29	1.33	-0.015	-0.024
	(1.41)	(1.35)	(0.015)	(0.018)
Suburb Districts	1.35	1.36	-0.009	0.005
	(1.13)	(1.18)	(0.011)	(0.018)
Town Districts	0.22	0.32	-0.011	0.046**
	(1.25)	(1.24)	(0.014)	(0.020)
Rural Districts	-0.89	-0.95	-0.011	-0.002
	(0.80)	(0.81)	(0.009)	(0.011)
Panel B: Grade 4 Reading				
Overall	-0.61	-0.43	-0.009*	0.003
	(0.66)	(0.51)	(0.005)	(0.007)
Majority-Black Districts	2.40	2.70	0.007	0.016
	(2.84)	(2.02)	(0.025)	(0.032)
Majority-Hispanic Districts	-5.12	-3.65	-0.015	0.030
	(3.56)	(2.47)	(0.025)	(0.035)
90+% White Districts	-0.33	-0.44	-0.006	-0.003
	(0.89)	(0.75)	(0.008)	(0.011)
40+% FRL Districts	-0.40	-0.19	-0.005	0.004
	(0.88)	(0.66)	(0.007)	(0.010)
<40% FRL Districts	-0.61	-0.50	-0.011	-0.004
	(0.90)	(0.77)	(0.007)	(0.010)
2500+ Student Districts	0.26	0.33	-0.013*	-0.003
	(1.00)	(0.77)	(0.008)	(0.009)
<2500 Student Districts	-1.56*	-1.31**	-0.005	0.008
	(0.80)	(0.66)	(0.007)	(0.010)
City Districts	-0.68	-0.34	-0.018	-0.010
	(1.98)	(1.46)	(0.014)	(0.016)
Suburb Districts	-0.61	-0.72	0.003	0.011
	(1.32)	(1.10)	(0.012)	(0.016)
Town Districts	-0.54	-0.20	-0.018	0.018
	(1.32)	(1.08)	(0.012)	(0.016)
Rural Districts	-0.22	-0.31	-0.001	-0.001
	(0.93)	(0.74)	(0.008)	(0.011)

Table 5 The Effects of Pre-K, Using District-Level Variation, cont'd

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel C: Grade 8 Math				
Overall	-0.78	-0.59	-0.002	0.012
	(0.77)	(0.59)	(0.006)	(0.010)
Majority-Black Districts	-4.11	-1.62	-0.021	-0.016
	(3.18)	(2.30)	(0.027)	(0.047)
Majority-Hispanic Districts	-9.87**	-7.67**	0.024	-0.050
	(4.87)	(3.70)	(0.023)	(0.048)
90+% White Districts	-0.74	-0.65	0.004	0.002
	(1.04)	(0.87)	(0.011)	(0.017)
40+% FRL Districts	-2.19**	-1.67**	-0.001	0.009
	(1.00)	(0.75)	(0.008)	(0.014)
<40% FRL Districts	0.53	0.34	-0.012	0.002
	(1.15)	(0.95)	(0.010)	(0.015)
2500+ Student Districts	-1.58	-1.08	-0.006	0.006
	(1.15)	(0.85)	(0.009)	(0.014)
<2500 Student Districts	0.97	0.62	0.009	0.022
	(0.97)	(0.80)	(0.009)	(0.015)
City Districts	-2.28	-1.49	-0.003	-0.004
	(2.00)	(1.53)	(0.017)	(0.019)
Suburb Districts	0.75	0.80	-0.017	-0.017
	(1.85)	(1.45)	(0.015)	(0.016)
Town Districts	1.24	0.85	0.021	0.014
	(1.78)	(1.44)	(0.015)	(0.016)
Rural Districts	0.15	-0.24	-0.001	-0.008
	(1.08)	(0.87)	(0.010)	(0.011)
Panel D: Grade 8 Reading				
Overall	-1.18	-1.03	-0.004	-0.001
	(0.74)	(0.62)	(0.006)	(0.010)
Majority-Black Districts	1.76	1.10	-0.044	0.019
	(4.29)	(3.28)	(0.029)	(0.034)
Majority-Hispanic Districts	-4.24	-3.45	-0.012	-0.070
	(3.82)	(3.32)	(0.027)	(0.051)
90+% White Districts	-1.17	-0.76	0.003	0.011
	(1.16)	(1.04)	(0.012)	(0.015)
40+% FRL Districts	-1.60	-1.34	-0.012	-0.006
	(1.03)	(0.84)	(0.009)	(0.014)
<40% FRL Districts	-0.72	-0.66	0.003	-0.015
	(1.11)	(1.02)	(0.010)	(0.013)
2500+ Student Districts	-2.03	-1.82**	-0.006	0.008
	(1.08)	(0.91)	(0.008)	(0.013)
<2500 Student Districts	0.72	0.54	-0.005	-0.014
	(0.97)	(0.86)	(0.010)	(0.013)
City Districts	0.34	-0.46	-0.011	-0.008
	(2.36)	(1.97)	(0.016)	(0.025)
Suburb Districts	-0.80	-0.69	0.002	0.003
	(1.52)	(1.32)	(0.013)	(0.018)
Town Districts	-2.87*	-2.35*	0.037**	0.010
	(1.68)	(1.41)	(0.016)	(0.022)
Rural Districts	-0.15	-0.19	-0.022**	-0.005
	(1.09)	(0.93)	(0.011)	(0.015)

NOTE: Each cell is from a separate regression of the outcome on the pre-K measure, district fixed effects, a set of state-by-year dummies, and time-varying categorical indicators for share of students on free or reduced-price lunch, district size, racial composition, and urbanicity. With the exception of urbanicity, these controls are of finer gradation than indicated by the subsamples in the rows, which are based on averages over the sample period. For number of observations, see Table 2. Standard errors in parentheses are clustered by district.

Table 6 The Effects of Pre-K, Using School-Level Variation

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Overall	0.13	0.20	0.002	0.003
	(0.44)	(0.44)	(0.005)	(0.007)
Majority-Black Schools	-0.18	0.29	-0.008	-0.017
	(1.49)	(1.16)	(0.012)	(0.018)
Majority-Hispanic Schools	2.35*	2.62**	0.020	0.000
	(1.26)	(1.12)	(0.017)	(0.019)
90+% White Schools	-0.55	-0.52	-0.005	0.017
	(0.80)	(0.83)	(0.009)	(0.011)
50+% FRL Schools	0.23	0.41	0.012*	0.002
	(0.60)	(0.55)	(0.006)	(0.010)
<50% FRL Schools	0.28	0.26	-0.007	0.004
	(0.63)	(0.66)	(0.007)	(0.009)
400+ Student Schools	1.08*	1.02*	0.001	0.001
	(0.60)	(0.59)	(0.006)	(0.009)
<400 Student Schools	-0.97	-0.76	0.001	0.009
	(0.64)	(0.61)	(0.007)	(0.009)
City Schools	0.75	0.94	-0.003	-0.018*
	(0.68)	(0.65)	(0.007)	(0.010)
Suburb Schools	1.56	1.56	0.020*	0.033*
	(1.12)	(1.10)	(0.011)	(0.017)
Town Schools	-1.85	-1.70	0.001	0.039*
	(1.58)	(1.56)	(0.014)	(0.021)
Rural Schools	-0.37	-0.50	-0.001	0.007
	(0.70)	(0.71)	(0.009)	(0.011)
Panel B: Grade 4 Reading				
Overall	0.39	-0.12	0.007	0.001
	(0.44)	(0.34)	(0.005)	(0.005)
Majority-Black Schools	-0.36	0.00	0.001	-0.003
	(1.19)	(0.85)	(0.011)	(0.012)
Majority-Hispanic Schools	0.34	-0.15	-0.012	0.008
	(1.67)	(1.13)	(0.016)	(0.016)
90+% White Schools	-0.41	-0.18	0.001	-0.004
	(0.73)	(0.63)	(0.008)	(0.010)
50+% FRL Schools	-0.01	0.19	0.011	0.010
	(0.66)	(0.46)	(0.007)	(0.007)
<50% FRL Schools	-0.90	-0.57	0.005	-0.009
	(0.58)	(0.49)	(0.006)	(0.008)
400+ Student Schools	-0.25	-0.01	0.010	-0.008
	(0.61)	(0.46)	(0.006)	(0.007)
<400 Student Schools	-0.70	-0.35	0.005	0.010
	(0.62)	(0.50)	(0.007)	(0.008)
City Schools	0.62	0.56	0.008	-0.001
	(0.81)	(0.57)	(0.008)	(0.007)
Suburb Schools	-0.02	0.39	0.017*	-0.001
	(1.05)	(0.87)	(0.009)	(0.013)
Town Schools	-1.94	-1.17	0.012	0.034*
	(1.23)	(1.04)	(0.014)	(0.020)
Rural Schools	-0.78	-0.57	0.002	-0.004
	(0.79)	(0.63)	(0.008)	(0.010)

NOTE: Each cell is from a separate regression of the outcome on the pre-K measure, school fixed effects, a set of state-by-year dummies, and time-varying categorical indicators for share of students on free or reduced-price lunch, school size, racial composition, and urbanicity. With the exception of urbanicity, these controls are of finer gradation than indicated by the subsamples in the rows, which are based on averages over the sample period. For number of observations, see Table 2. Standard errors in parentheses are clustered by district.

Table 7 The Effects of Pre-K by Race, Using State-Level Variation and NIEER Quality Measures

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Whites * Quartile 1	0.60	0.48	0.013	0.002
	(5.32)	(5.72)	(0.027)	(0.055)
Whites * Quartile 2	-0.96	-0.83	0.015	-0.073
	(3.78)	(4.14)	(0.028)	(0.064)
Whites * Quartile 3	-1.70	-1.18	0.015	-0.089
	(4.21)	(4.67)	(0.027)	(0.090)
Whites * Quartile 4	-0.93	-0.35	0.037	-0.066
	(3.92)	(4.26)	(0.028)	(0.087)
Blacks * Quartile 1	7.23*	5.04	-0.015	0.213**
	(3.45)	(3.06)	(0.034)	(0.101)
Blacks * Quartile 2	4.46	3.11	-0.016	-0.016
	(3.84)	(3.25)	(0.028)	(0.056)
Blacks * Quartile 3	2.54	1.27	-0.025	0.023
	(3.51)	(2.94)	(0.024)	(0.072)
Blacks * Quartile 4	1.03	-0.24	-0.016	0.025
	(3.63)	(3.06)	(0.027)	(0.080)
Hispanics * Quartile 1	2.82	3.38	0.030	-0.068
	(7.11)	(6.56)	(0.050)	(0.125)
Hispanics * Quartile 2	1.58	1.67	0.051	-0.105
	(6.01)	(5.27)	(0.032)	(0.136)
Hispanics * Quartile 3	-0.00	0.32	0.043	-0.195
	(7.49)	(6.69)	(0.038)	(0.146)
Hispanics * Quartile 4	-3.28	-1.88	0.117**	-0.193
	(8.04)	(7.13)	(0.048)	(0.139)
Others * Quartile 1	6.00	6.42	0.002	0.094
	(10.78)	(9.49)	(0.059)	(0.156)
Others * Quartile 2	7.96	8.80	0.081	0.100
	(12.91)	(11.74)	(0.069)	(0.256)
Others * Quartile 3	3.97	3.15	0.119	-0.097
	(13.14)	(11.90)	(0.080)	(0.209)
Others * Quartile 4	2.27	1.67	0.118*	-0.086
	(10.77)	(9.77)	(0.067)	(0.181)

Table 7 The Effects of Pre-K by Race, Using State-Level Variation and NIEER Quality Measures, cont'd

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel B: Grade 4 Reading				
Whites * Quartile 1	0.02	0.12	0.004	-0.032
	(4.01)	(3.59)	(0.033)	(0.049)
Whites * Quartile 2	-0.50	-0.36	0.003	-0.027
	(3.59)	(3.13)	(0.022)	(0.066)
Whites * Quartile 3	0.67	0.07	-0.012	-0.055
	(3.67)	(3.17)	(0.024)	(0.087)
Whites * Quartile 4	-0.43	-0.62	0.021	-0.037
	(3.65)	(3.08)	(0.024)	(0.085)
Blacks * Quartile 1	5.51	2.69	-0.058	0.114
	(4.89)	(3.36)	(0.048)	(0.104)
Blacks * Quartile 2	5.62	3.20	-0.043	-0.017
	(4.86)	(3.51)	(0.031)	(0.070)
Blacks * Quartile 3	5.88	2.23	-0.048	-0.068
	(4.94)	(3.39)	(0.033)	(0.086)
Blacks * Quartile 4	4.60	1.11	-0.007	-0.054
	(4.79)	(3.31)	(0.041)	(0.088)
Hispanics * Quartile 1	5.47	2.82	0.058	0.000
	(10.30)	(6.28)	(0.062)	(0.115)
Hispanics * Quartile 2	2.65	0.86	0.047	-0.028
	(8.06)	(5.51)	(0.051)	(0.117)
Hispanics * Quartile 3	5.05	1.90	0.061	-0.050
	(10.14)	(6.86)	(0.054)	(0.140)
Hispanics * Quartile 4	5.44	3.13	0.058	-0.062
	(11.54)	(7.42)	(0.055)	(0.145)
Others * Quartile 1	-4.06	-0.80	-0.043	-0.088
	(14.27)	(9.51)	(0.057)	(0.170)
Others * Quartile 2	-3.29	-2.22	-0.080	-0.179
	(12.27)	(9.19)	(0.056)	(0.174)
Others * Quartile 3	-3.04	-2.89	-0.047	-0.310*
	(13.62)	(9.81)	(0.062)	(0.181)
Others * Quartile 4	-2.13	-4.13	-0.033	-0.229
	(11.16)	(7.63)	(0.051)	(0.160)

NOTE: Each set of quartiles and race is from a separate regression of the outcome on the (race-specific) pre-K measure interacted with NIEER quality quartiles (using PCA), a set of state dummies, and a set of test year dummies. State-year cells that lacked NIEER quality metrics were interacted with a separate dummy, the coefficient for which is not shown. Each observation is a state-year, and there are 193 observations for each regression, except for column 4, in which there are 144 observations. Standard errors in parentheses are clustered by state.

Data Appendix

Pre-K

The Common Core of Data (CCD) provides enrollment for the universe of public schools in the United States. For state-level analyses, we take reported prekindergarten, both overall and by race, from the CCD's state-level files and divide by state-year estimates of the population of 4-year-olds from National Cancer Institute's SEER population data. For school-level analyses, we take reported prekindergarten and grade 1 enrollment from the CCD's school-level files, and divide the former by the latter, topcoding the ratio at 1 if it exceeds 1 but is less than 1.5 and set to missing ratios that exceed 1.5. For district-level analyses, we again take reported prekindergarten and grade 1 enrollment from the CCD's school-level files, as grade-specific enrollment is not reported in the district-level files. We sum enrollments in each grade for all schools within a district, and then take the ratios of these sums, with the same topcoding rule applied. Pre-K enrollment by school and race is available in recent years, but we do not use it given its limited availability.

Not every school or state reports a valid number for pre-K enrollment each year. In most of these cases, there is a missing code for not applicable. That is, instead of entering a zero, the school or district reporting official indicated that the pre-K enrollment field was not applicable because there was no pre-K program. In some other cases, on the other hand, it appears that the state or school may have pre-K enrollment but report pre-K enrollment as a true missing (different than the "not applicable: missing code). California, for example, never reports pre-K enrollment by school or by race for the state as a whole, but does report positive pre-K enrollment for the state in the aggregate. We code the "not applicable" missings as true zeros and the true missings as such, with the following exceptions: (1) if a school or state reports positive pre-K enrollment in year *t-1* and year *t+1* but a "not applicable in year *t*, we code it as a missing; (2) if positive pre-K was reported at the state level but no school in that state and year reported positive pre-K enrollment, all such schools were coded to missing that year.

Appendix Table 1: A Summary of the Literature of the Effects of Pre-K Programs Over Various Horizons

Type of study	Study	Short-run (< 1 year)	Medium-run (primary school)	Long-run (high school+)
Classic experiments	Perry: 2 yrs of half-day pre-K, @$10,427 per student-yr.	18 percentiles (ES=0.59)	3 percentiles at end of 3rd grade (ES=0.10), 1 percentile at end of 4th grade (ES=0.04). Reduces special ed for mental impairment by 20 pp, overall special ed by 5 pp. (ns). Reduces grade repetition by 5 pp, grade repetition by 2 or more years by 7 pp (ns).	19% earnings boost; 50–59% crime reduction; reduced smoking/drinking
	Abecedarian: 5 yrs of full-time full-yr care/pre-K, birth to 5, @$17,633 per student-yr	19 percentiles (ES=0.50)	10 percentiles at 3rd grade (ES=0.27)	26% earnings boost; no crime effect; reduced risk factors for cardiovascular disease
Quasi-experiments	Chicago Child-Parent Center: 2 yrs of half-day pre-K, @$5,597 per student-yr. Benefits did not increase much for 2-yr vs. 1-yr.	11 percentiles (ES=0.38)	3 percentiles at 3rd grade (ES=0.07); grade retention by age 15 drops by 15pp; special ed by age 18 drops by 10pp.	8% earnings boost; 22% reduction in felony arrests; 26% reduction in depression, 24% reduction in substance abuse.
	Head Start-siblings (Deming): 1–2 yrs of mix of half-day versus full-day, although modal is 1-year, @$9,134 per student-year.	5 percentiles at ages 5-6 (ES=0.15)	4 percentiles at ages 7-10 (ES=0.13); 2 percentiles at ages 11-14 (ES=0.06). Reduced diagnosis of learning disability by 6 pp, ever grade repetition by 7 pp.	Predicted 11% earnings gain; no crime effect; percentage in poor health drop by 7 pp.
	Head Start-siblings (Currie & Thomas; Garces, Thomas, & Currie)	Currie & Thomas: 7 percentiles at age 5 (ES=0.21)	Currie and Thomas: 6 percentiles (ES=0.18) for whites, 0 for blacks. White reduction in any grade retention by age 10+ is 47pp, 0 reduction for blacks.	Garces-Currie-Thomas: whites 28 pps more likely to complete high school, 28 pps more likely to attend college; 0 attainment effects for blacks. Blacks 13 pp less likely to be charged with crime, no white effects.
	Head Start (Ludwig & Miller) comparison across counties with different grantwriting assistance (geographic study).		Grantwriting assistance reduces Head Start preventable mortality at ages 5–9 by 30–50 percent. No effects on 8th grade test scores.	Grantwriting assistance increases high school completion and college attendance by 3 to 5 pp.
Meta-analyses	(Duncan & Magnuson)	9 percentiles at end of program (ES=.27)	5 percentiles by 4th grade (ES=.15)	

	Camilli, Barnett et al.	14 percentiles at end of program (ES=.39)	4 to 5 percentiles both at ages 5–10 and ages 10+ (ES=0.14–0.15)	
Other studies	Head Start Experiment	7 percentiles at end of program (ES=0.22)	2 percentiles at end of 3rd grade (ES=0.06)	
	RDD Barnett et al. studies of 7 states	11 percentiles at beginning of kindergarten (ES=0.31)		
	RDD Gormley et al. (Tulsa) and matching follow-up study. $5,238 for half-day pre-K for one school year, $10,476 for full-day pre-K.	RDD results: At kdg entrance, full-day has pctile gain of 18 for FRL students, 17 for non; half-day is 11 for FRL, 10 for non (ESs = 1.07, 0.96, 0.66, 0.58) . PSM results appear to cut these ESs in half for reading, by 1/3rd for math.	7 percentiles (ES=0.18) in math for late cohort, less than 0.4 pctiles (ES=0.01) for early cohort in math. In reading, 4 percentiles for late cohort (ES =0.09), minus 1 percentile for early cohort (ES=-0.03) . Only late cohort math result is statistically significant.	
	RDD Weiland/Yoshikawa (Boston). Full-day pre-K program, cost of $15,000 to $17,000 per student.	21 percentiles gain at kindergarten entrance for FRL students (ES =.59), 15 percentiles for non-FRL students (ES= .38)		
	Tennessee experiment (Lipsey et al.) Full-day 1-year program at $4,611 per student.	8 percentile gain at end of program (ES=0.24) based on comparison group. 17 percentile gain at kindergarten entry (ES=0.49) based on RDD.	3 percentile LOSS at end of 3rd grade (ES=-0.1).	
Kindergarten class quality	Chetty et al.: 1 standard deviation improvement in kindergarten class quality, as measured by end of kindergarten peer scores.	6 percentile gain at end of kindergarten (ES=0.16)	1 percentile gain at end of 4th grade (ES=0.03)	3% gain in adult earnings
Recent geographic studies	Fitzpatrick (Georgia): Georgia: $5,520 per student for full-day program.		6 percentile points (ES=.15) for both math and reading NAEP scores at 4th grade; significant with clustered standard errors, insignificant with Conley-Table corrections.	
	Cascio/Schanzenbach (Oklahoma/Georgia): OK: $7,685 per student for mix of half-day and full-day programs: GA: $5,520 per		4th grade: FRL gain of 14 percentiles in both math & reading NAEP scores (ES=.39, .40); non-FRL gain of 4 pctiles in math, loss of 6 pctiles in	

	student for full-day program.		reading (ES=.10, -.16). 8th grade: FRL gain of 11 pctiles in math, 4 pctiles in reading (ES=.33, .12); non-FRL loss of 5 pctiles in math, 4 pctiles in reading (ES=-.12, -.09). Only FRL 4th grade gains and 8th grade math gains are statistically significant in main reported estimates; none of estimates are statistically significant with Conley-Taber corrections.	
	Ladd et al. (North Carolina) More at Four, a full-day pre-K program, @$6,066 per student.		20 percentiles in math (ES=.54), 25 percentiles in reading for North Carolina tests (ES=.66)	
	Rosinsky, panel data on all states		State funded pre-K reduces 4th grade math NAEP test scores by 6 percentiles for all students (ES=-.14), and 7 percentiles for low-income students (ES=-.26). All publicly funded pre-K reduces NAEP scores of all students by 5 percentiles (ES=-.11), low-income students by 6 percentiles (ES=-.20).	

Appendix Table 2: States and Years with Math NAEP data and Valid Pre-K Measures

State FIPS code	1996	1998	2000	2002	2003	2005	2007	2009	2011	2013
Alabama								X	X	X
Alaska					X	X	X	X	X	X
Arizona			X		X	X	X	X	X	X
Arkansas			X		X	X	X	X	X	X
California										
Colorado					X	X	X	X	X	X
Connecticut			X		X	X	X	X	X	X
Delaware					X	X	X	X	X	X
District of Columbia			X		X	X	X	X	X	X
Florida					X	X	X	X	X	X
Georgia			X		X	X	X	X	X	X
Hawaii			X		X	X	X	X	X	X
Idaho								X	X	X
Illinois			X		X	X	X	X	X	X
Indiana			X		X	X	X	X	X	X
Iowa			X		X	X	X	X	X	X
Kansas			X			X	X	X	X	X
Kentucky						X	X			X
Louisiana			X		X	X	X	X	X	X
Maine			X		X	X	X	X	X	X
Maryland			X		X	X	X	X	X	X
Massachusetts			X		X	X	X	X	X	X
Michigan			X		X	X	X	X	X	X
Minnesota			X		X	X	X	X	X	X
Mississippi			X		X	X	X	X	X	X
Missouri			X		X	X	X	X	X	X
Montana			X		X	X	X	X	X	X
Nebraska			X		X	X	X	X	X	X
Nevada			X		X	X	X	X	X	X
New Hampshire					X	X	X	X	X	X
New Jersey							X	X	X	X
New Mexico			X		X	X	X	X	X	X
New York			X		X	X	X	X	X	X
North Carolina			X			X	X	X	X	X
North Dakota					X	X	X	X	X	X
Ohio			X		X	X	X	X	X	X
Oklahoma			X		X	X	X	X	X	X
Oregon			X		X	X	X	X	X	X
Pennsylvania					X	X	X	X	X	X
Rhode Island			X		X	X	X	X	X	X
South Carolina					X	X	X	X	X	X
South Dakota					X	X	X	X	X	X
Tennessee									X	X
Texas			X		X	X	X	X	X	X
Utah			X		X	X	X	X	X	X
Vermont			X		X	X	X	X	X	X
Virginia			X		X	X	X	X	X	X
Washington					X	X	X	X	X	X
West Virginia			X		X	X	X	X	X	X
Wisconsin			X		X	X	X	X	X	X
Wyoming						X		X	X	X

NOTE: Pre-K data is lagged five years from shown (NAEP) year. The Math NAEP was not conducted in 1998 and 2002.

Appendix Table 3: States and Years with Reading NAEP data and Valid Pre-K Measures

State FIPS code	1996	1998	2000	2002	2003	2005	2007	2009	2011	2013
Alabama								X	X	X
Alaska				X	X	X	X	X	X	X
Arizona		X		X	X	X	X	X	X	X
Arkansas		X		X	X	X	X	X	X	X
California										
Colorado		X			X	X	X	X	X	X
Connecticut		X		X	X	X	X	X	X	X
Delaware		X		X	X	X	X	X	X	X
District of Columbia		X		X	X	X	X	X	X	X
Florida		X		X	X	X	X	X	X	X
Georgia		X		X	X	X	X	X	X	X
Hawaii		X		X	X	X	X	X	X	X
Idaho								X	X	X
Illinois		X		X	X	X	X	X	X	X
Indiana				X	X	X	X	X	X	X
Iowa		X		X	X	X	X	X	X	X
Kansas		X		X		X	X	X	X	X
Kentucky						X	X			X
Louisiana		X		X	X	X	X	X	X	X
Maine				X	X	X	X	X	X	X
Maryland		X		X	X	X	X	X	X	X
Massachusetts		X		X	X	X	X	X	X	X
Michigan		X		X	X	X	X	X	X	X
Minnesota		X		X	X	X	X	X	X	X
Mississippi		X		X	X	X	X	X	X	X
Missouri				X	X	X	X	X	X	X
Montana		X		X	X	X	X	X	X	X
Nebraska				X	X	X	X	X	X	X
Nevada		X		X	X	X	X	X	X	X
New Hampshire		X		X	X	X	X	X	X	X
New Jersey							X	X	X	X
New Mexico		X		X	X	X	X	X	X	X
New York		X		X	X	X	X	X	X	X
North Carolina		X		X		X	X	X	X	X
North Dakota				X	X	X	X	X	X	X
Ohio				X	X	X	X	X	X	X
Oklahoma		X		X	X	X	X	X	X	X
Oregon		X		X	X	X	X	X	X	X
Pennsylvania				X	X	X	X	X	X	X
Rhode Island		X		X	X	X	X	X	X	X
South Carolina						X	X	X	X	X
South Dakota				X	X	X	X	X	X	X
Tennessee									X	X
Texas		X		X	X	X	X	X	X	X
Utah		X		X	X	X	X	X	X	X
Vermont				X	X	X	X	X	X	X
Virginia		X		X	X	X	X	X	X	X
Washington		X		X	X	X	X	X	X	X
West Virginia		X		X	X	X	X	X	X	X
Wisconsin		X		X	X	X	X	X	X	X
Wyoming						X		X	X	X

NOTE: Pre-K data is lagged five years from shown (NAEP) year. The Reading NAEP was not conducted in 1996 and 2000.

Appendix Table 4 The Effects of Pre-K by Race, Using State-Level Variation, Including State-Linear Trends

	(1) Scale score	(2) Percentile Score	(3) IEP	(4) Overage
Panel A: Grade 4 Math				
Whites	-3.63	-3.80	0.008	-0.034
	(3.23)	(3.32)	(0.027)	(0.027)
Blacks	6.10	6.07	-0.011	0.090**
	(3.86)	(4.01)	(0.027)	(0.040)
Hispanics	-0.96	0.70	-0.037	-0.028
	(3.30)	(2.86)	(0.030)	(0.088)
Others	1.83	1.69	0.018	-0.008
	(3.03)	(2.93)	(0.050)	(0.053)
Panel B: Grade 4 Reading				
Whites	-1.75	-2.12	-0.009	-0.009
	(3.03)	(2.45)	(0.017)	(0.039)
Blacks	5.60**	4.47**	-0.023	0.036
	(1.76)	(1.85)	(0.016)	(0.059)
Hispanics	0.41	0.15	-0.034	0.005
	(4.17)	(2.58)	(0.051)	(0.066)
Others	-0.53	-0.25	0.025	-0.018
	(5.08)	(3.29)	(0.022)	(0.051)
Panel C: Grade 8 Math				
Whites	3.17	3.06	-0.062	-0.053
	(3.62)	(3.14)	(0.044)	(0.065)
Blacks	1.74	-0.08	0.027	-0.056
	(4.65)	(3.13)	(0.028)	(0.051)
Hispanics	-1.54	-3.34	-0.047	-0.097
	(6.73)	(6.33)	(0.060)	(0.087)
Others	-8.90	-7.29	0.038	-0.055
	(6.14)	(5.05)	(0.028)	(0.068)
Panel D: Grade 8 Reading				
Whites	-1.31	-1.27	-0.053	0.027
	(2.92)	(2.31)	(0.039)	(0.043)
Blacks	1.92	0.07	0.004	-0.003
	(2.61)	(2.17)	(0.030)	(0.068)
Hispanics	2.68	0.53	-0.040	-0.095
	(8.77)	(6.74)	(0.044)	(0.095)
Others	-4.64	-2.22	0.032*	-0.031
	(4.39)	(3.96)	(0.018)	(0.049)

NOTE: Each cell is from a separate regression of the outcome on the (race-specific) pre-K measure, a set of state dummies, a set of test year dummies, and a set of state-specific linear time trends. For other notes, see Table 3.

Appendix Table 5 The Effects of Pre-K by School and Student Income, Using State-Level Variation, Including State-Linear Trends

	(1) Scale score	(2) Percentile Score	(3) IEP	(4) Overage
Panel A: Grade 4 Math				
Low-Income Schools	2.62	3.41	-0.003	0.016
	(3.22)	(3.09)	(0.030)	(0.035)
Non-low-income Schools	2.58	2.57	-0.008	0.036
	(3.10)	(3.18)	(0.016)	(0.024)
Low-Income Students	1.84	2.41	-0.003	0.022
	(2.37)	(2.64)	(0.023)	(0.031)
Non-low-income Students	3.89	3.31	0.014	-0.003
	(3.76)	(3.79)	(0.016)	(0.029)
Panel B: Grade 4 Reading				
Low-Income Schools	1.38	0.54	-0.007	0.030
	(4.09)	(2.79)	(0.022)	(0.048)
Non-low-income Schools	0.10	0.19	0.006	0.061
	(2.25)	(1.82)	(0.015)	(0.039)
Low-Income Students	1.76	1.13	0.005	0.051
	(2.75)	(1.83)	(0.026)	(0.041)
Non-low-income Students	1.14	0.69	0.002	0.020
	(3.21)	(2.84)	(0.016)	(0.031)
Panel C: Grade 8 Math				
Low-Income Schools	5.16	2.76	0.002	-0.093
	(3.49)	(2.52)	(0.028)	(0.059)
Non-low-income Schools	-0.70	-0.22	-0.042	0.007
	(4.24)	(3.61)	(0.039)	(0.051)
Low-Income Students	1.18	0.16	-0.020	-0.038
	(3.85)	(3.17)	(0.024)	(0.072)
Non-low-income Students	-1.29	-0.36	-0.017	-0.070
	(4.09)	(3.26)	(0.021)	(0.079)
Panel D: Grade 8 Reading				
Low-Income Schools	2.57	0.67	0.022	-0.006
	(2.95)	(1.93)	(0.036)	(0.043)
Non-low-income Schools	-0.41	-0.34	-0.075**	0.082**
	(3.48)	(2.95)	(0.037)	(0.034)
Low-Income Students	1.16	-0.23	0.001	-0.006
	(3.60)	(2.53)	(0.036)	(0.037)
Non-low-income Students	-1.81	-1.60	-0.029	-0.032
	(2.74)	(2.38)	(0.023)	(0.044)

NOTE: Each cell is from a separate regression of the outcome on the pre-K measure, a set of state dummies, a set of test year dummies, and a set of state-specific linear time trends. For other notes, see Table 4.

Appendix Table 6 The Effects of Pre-K by Race, Using State-Level Variation and NIEER Spending Measures

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel A: Grade 4 Math				
Whites * Quartile 1	-3.32	-3.23	0.035	-0.016
	(3.46)	(3.76)	(0.027)	(0.038)
Whites * Quartile 2	1.94	1.99	0.071**	0.003
	(4.04)	(4.39)	(0.029)	(0.061)
Whites * Quartile 3	3.04	3.41	0.018	-0.064
	(3.49)	(3.92)	(0.024)	(0.055)
Whites * Quartile 4	-1.06	-0.72	0.050*	-0.012
	(3.58)	(3.89)	(0.027)	(0.051)
Blacks * Quartile 1	11.89**	8.94**	-0.019	0.055
	(4.50)	(3.75)	(0.018)	(0.057)
Blacks * Quartile 2	11.63**	8.66**	-0.017	0.040
	(5.14)	(4.25)	(0.023)	(0.070)
Blacks * Quartile 3	9.54*	6.92	0.002	0.015
	(5.11)	(4.32)	(0.019)	(0.083)
Blacks * Quartile 4	9.05**	7.25**	0.013	0.054
	(3.83)	(3.27)	(0.011)	(0.053)
Hispanics * Quartile 1	6.06	5.63	-0.001	-0.060
	(4.19)	(3.45)	(0.032)	(0.125)
Hispanics * Quartile 2	7.03	6.41*	0.012	-0.112
	(4.32)	(3.58)	(0.035)	(0.110)
Hispanics * Quartile 3	11.26*	10.02*	0.016	-0.026
	(6.25)	(5.78)	(0.039)	(0.126)
Hispanics * Quartile 4	3.90	3.43	0.045	-0.033
	(3.47)	(2.70)	(0.034)	(0.102)
Others * Quartile 1	12.83	11.71	0.030	-0.035
	(7.91)	(7.44)	(0.064)	(0.115)
Others * Quartile 2	13.69	12.95	0.058	-0.126
	(8.74)	(8.14)	(0.064)	(0.130)
Others * Quartile 3	16.15	14.85	0.029	-0.138
	(10.01)	(9.59)	(0.061)	(0.129)
Others * Quartile 4	8.56	7.91	0.046	-0.005
	(6.93)	(6.56)	(0.049)	(0.099)

Appendix Table 6, cont'd

	(1)	(2)	(3)	(4)
	Scale score	Percentile Score	IEP	Overage
Panel B: Grade 4 Reading				
Whites * Quartile 1	0.44	0.10	0.007	-0.020
	(3.98)	(3.38)	(0.026)	(0.041)
Whites * Quartile 2	1.99	1.82	0.006	-0.049
	(4.09)	(3.60)	(0.030)	(0.055)
Whites * Quartile 3	-0.01	-0.09	0.005	-0.038
	(3.60)	(3.12)	(0.028)	(0.053)
Whites * Quartile 4	-0.21	-0.39	0.013	-0.014
	(4.00)	(3.30)	(0.026)	(0.047)
Blacks * Quartile 1	8.78**	6.10*	-0.037*	0.060
	(4.23)	(3.06)	(0.021)	(0.080)
Blacks * Quartile 2	5.94	4.31	-0.019	0.036
	(4.53)	(3.30)	(0.024)	(0.086)
Blacks * Quartile 3	4.36	2.66	-0.020	0.013
	(4.41)	(3.13)	(0.021)	(0.077)
Blacks * Quartile 4	4.82*	3.74	-0.001	0.005
	(2.84)	(2.31)	(0.011)	(0.058)
Hispanics * Quartile 1	0.69	-0.50	0.047	0.023
	(5.63)	(3.99)	(0.041)	(0.113)
Hispanics * Quartile 2	-1.62	-1.56	0.068	0.036
	(5.72)	(4.07)	(0.041)	(0.104)
Hispanics * Quartile 3	-1.06*	-1.21	0.047	0.156
	(6.49)	(4.57)	(0.041)	(0.113)
Hispanics * Quartile 4	0.79	0.15	0.033	0.037
	(4.62)	(3.24)	(0.034)	(0.102)
Others * Quartile 1	3.59	1.77	-0.047	-0.027
	(7.91)	(7.50)	(0.055)	(0.174)
Others * Quartile 2	-0.17	-0.97	0.002	-0.021
	(11.64)	(8.29)	(0.062)	(0.181)
Others * Quartile 3	2.41	1.43	-0.027	-0.067
	(10.78)	(7.83)	(0.060)	(0.178)
Others * Quartile 4	2.19	0.69	-0.035	-0.015
	(9.25)	(6.38)	(0.044)	(0.147)

NOTE: Each set of quartiles and race is from a separate regression of the outcome on the (race-specific) pre-K measure interacted with NIEER spending quartiles, a set of state dummies, and a set of test year dummies. State-year cells that lacked NIEER quality metrics were interacted with a separate dummy, the coefficient for which is not shown. Each observation is a state-year, and there are 193 observations for each regression, except for column 4, in which there are 144 observations. Standard errors in parentheses are clustered by state.