
Does Universal Preschool Hit the Target?

Program Access and Preschool Impacts

Elizabeth U. Cascio

ABSTRACT


This study examines the cost efficacy of universal programs, taking advantage of the rich diversity in rules governing access to state-funded preschool in the United States. Using age-eligibility rules for identification, I find that attending a state-funded universal preschool generates substantial immediate test score gains, particularly for low-income children. Gains for low-income children from attending targeted (largely means-tested) preschool are significantly smaller. Cross-state differences in alternative care options, demographics, and other program features cannot explain the difference in attendance impacts across program types. Benefit-to-cost ratios of universal programs are favorable despite their relatively high costs per low-income child.

Elizabeth U. Cascio is a Professor of Economics at Dartmouth College (elizabeth.u.cascio@dartmouth.edu). For their helpful comments, the author thanks seminar and lecture participants at Aarhus University, American University, Boston College, the Federal Reserve Bank of Boston, Franklin and Marshall College, the Institute for Fiscal Studies, McMaster University, Montana State University, Ohio State University, Southern Methodist University, the University of Connecticut, the University of Southern Denmark, the 8th Annual International Workshop on Applied Economics of Education, and the 2018 AEA Annual Meeting. The author has no relevant interests to disclose. IRB approval was not sought because the analysis relies on secondary data. Nevertheless, to protect and ensure the confidentiality of respondents, the Institute of Education Sciences (IES) in the U.S. Department of Education requires users of the data to have a Restricted-Use Data License and to submit any papers or presentations using the data to the IES Data Security Office for review for disclosure risk prior to their circulation to non-licensed parties. The author has an IES Restricted-Use Data License, and no disclosure risks were found upon the final review of this article. While the author cannot share the data, they can be obtained via the application process described at <https://nces.ed.gov/statprog/rudman/>. The author is willing to assist in this process and to the extent possible will share code used to generate samples and results.

[Submitted February 2020; accepted November 2020]; doi:10.3368/jhr.58.3.0220-10728R1

JEL Classification: H75, I24, I28, J13, and J24

ISSN 0022-166X E-ISSN 1548-8004 © 2023 by the Board of Regents of the University of Wisconsin System

 Supplementary materials are freely available online at: <http://uwpress.wisc.edu/journals/journals/jhr-supplementary.html>

Elizabeth U. Cascio <https://orcid.org/0000-0002-2912-9970>

I. Introduction

In the context of many public programs, key policy parameters involve not just how but *whom*—which populations should be eligible for benefits. This is evident in recent policy proposals at the federal level in the United States, which depart from the targeting characteristic of much U.S. policy to expand benefits to all individuals who meet broader eligibility criteria, even when many could be inframarginal for the good or service in question. Central to the rhetoric behind such proposals, such as Medicare for All or universal childcare, is that access to some services is a basic right, not a privilege for those who can afford it. Another rationale is that “programs for the poor are poor programs”—that means-tested programs end up underfunded due to lack of broad-based political support.

But can universal programs ever be justified on efficiency grounds? In theory, there are clear efficiency rationales for targeting: by targeting benefits on difficult-to-change characteristics correlated with low levels of human capital, policymakers can reduce moral hazard, keep costs down, and redistribute toward those most in need. On the other hand, the greater number—and political power—of stakeholders in universal programs might hold public goods providers more accountable, raising the productivity of public spending relative to a targeted program. Particularly in cases where there are direct interactions across program beneficiaries, universal access may also allow for human capital spillovers that increase the productivity of public spending for an implicit target population, such as disadvantaged children.

It is nevertheless difficult to gain empirical traction on this question; in the very least, it requires that the same kind of program be observed under widely different conditions of access, a situation that rarely arises in practice. U.S. preschool education may provide a rare proving ground. Perhaps nowhere today is the variation in program access more striking: in 2015–2016, not all states funded prekindergarten (pre-K) programs, and among the 43 states that did, there was great cross-state variation in eligibility rules. Some state programs are universal, serving all four-year-olds that meet age-eligibility requirements; others are targeted, meaning that they are also means-tested or target enrollment based on other risk factors (Barnett et al. 2017).

In the first part of this paper, I take advantage of this rich cross-state variation in rules governing eligibility for state-funded pre-K to compare universal and targeted programs on the same basis, using the same data and research design. Despite a large literature on preschool education spanning disciplines, such an exercise has yet to be carried out. A mature body of research explores targeted preschool programs, most famously the federal Head Start program and the “model” preschool interventions of the distant past.¹ There is also emergent research on state-funded universal pre-K.² Both streams of literature tend to conclude that the benefits of preschool exceed the costs. Yet it is difficult to

1. Studies on Head Start include Currie and Thomas (1995); Garces, Thomas, and Currie (2002); Ludwig and Miller (2007); Deming (2009); Puma et al. (2010); Aizer and Cunha (2012); Carneiro and Ginja (2014); Bitler, Hoynes, and Domina (2014); Walters (2015); Kline and Walters (2016); Thompson (2018); Barr and Gibbs (2022); Anders, Barr, and Smith (2021); and Johnson and Jackson (2019). Regarding “model” interventions, see Heckman et al. (2010), Schweinhart et al. (2005), and recent reviews by Elango et al. (2016) and Almond, Currie, and Duque (2018). There are also studies on targeted state pre-K in North Carolina (Ladd, Muschkin, and Dodge 2014) and Tennessee (Lipsey et al. 2013; Lipsey, Farran, and Hofer 2016).

2. See, for example, Gormley and Gayer (2005), Fitzpatrick (2008), Cascio and Schanzenbach (2013), and Weiland and Yoshikawa (2013).

compare the findings for universal and targeted preschool programs directly due to differences across studies in methodology, outcomes, timing, and counterfactual enrollment patterns.³

I address this gap in the literature by working with survey data—the 2001 Birth Cohort of the Early Childhood Longitudinal Study (ECLS-B)—that span states where state-funded pre-K programs have different eligibility requirements and allow for credible estimation of the immediate gains from participation in universal and targeted programs alike. To estimate these gains, I take advantage of the large differences in state-funded pre-K eligibility and attendance among four-year-olds that arise due to state rules governing minimum age at school entry. The pre-K evaluation literature features a number of studies exploiting this variation using a regression discontinuity (RD) design.⁴ By contrast, I take a difference-in-differences (DD) approach, exploiting the larger gap in pre-K attendance rates of four-year-olds across adjacent school-entry cohorts in states with more robust state-funded pre-K programs. I thus use a comparison group to account for the direct effects of age on outcomes. The rich background characteristics available in the ECLS-B, including pretests, allow for useful tests of internal validity.

I find substantial positive effects of state-funded pre-K on the test scores of four-year-olds in states with universal programs: universal pre-K eligibility (attendance) improves the average four-year-old's standardized reading and math score by a significant 12 percent (60 percent) of a standard deviation. I cannot rule out equal effects of universal pre-K attendance by family income, but low-income children experience substantially larger test score gains.⁵ Pre-K attendance impacts for children in states with targeted programs are significantly smaller. Effect sizes vary, but this basic set of results is robust to changes in the estimation sample, including changing the age range or the set of states considered. It also arises for another outcome—parent reports of their four-year-old's kindergarten readiness—albeit less precisely. Supporting a causal interpretation, my preferred specification also does not yield similar patterns of impacts for a measure of mental development at age two or a host of other child observables.

A key feature of the universal and targeted pre-K programs under study is that the per-pupil costs are actually quite similar. The larger test score gains for universal programs alone thus suggest they are more cost-effective. But this conclusion could be hasty since some other characteristic of states with universal programs—rather than program access itself—could generate the same pattern of attendance impacts. For example, universal programs were relatively more likely to require small class sizes, which have been rigorously shown to improve early test performance (Krueger 1999). States with universal programs may also have populations more likely to benefit from formal preschool in general, or universal programs may be more likely to draw their

3. Wong et al. (2008) estimate the short-term cognitive effects of pre-K attendance in 2004–2005 in five states. Barnett et al. (2018) do a similar exercise in a more recent year for eight states. While the states differ some in terms of program access, these studies neither perform a formal analysis of the influence of access nor present estimates by family socioeconomic status.

4. The pre-K evaluation literature has looked to age-eligibility thresholds as a source of identifying variation since Gormley and Gayer's (2005) pioneering application of the RD design to Tulsa's pre-K program. See also, for example, Wong et al. (2008) and Weiland and Yoshikawa (2013). However, these RD studies have relied on district or state administrative data on public school *students*, precluding consistent estimates of pre-K eligibility and attendance impacts (Lipsey et al. 2015).

5. I define low income as eligibility for free or reduced-price lunch, since this is the modal income requirement for targeted programs. The substantive conclusions are robust to alternative definitions.

enrollees from informal or parental care. I rigorously explore these possibilities in the second part of the paper and find little supporting evidence. I also show that, consistent with access as a causal mechanism, the impacts of universal pre-K look quantitatively similar to those of universal public kindergarten within the ECLS-B, which I estimate by exploiting age-eligibility rules for kindergarten entry using an RD approach.

The evidence is thus consistent with universal pre-K being relatively cost-effective. In the third and final part of the paper, I monetize the test score gains in a tentative cost-benefit analysis. Even under conservative assumptions on key parameters like the magnitude of the association between early life test scores and earnings, universal pre-K delivers a benefit-to-cost ratio well above one, much like universal kindergarten. While less precise than the difference in test score effects of universal and targeted pre-K for low-income four-year-olds, the magnitude of the difference in benefit-to-cost ratios across programs is substantial.

The next section describes the landscape of state-funded preschool in the United States and the state-funded pre-K programs of study in this paper. Section III outlines the research design I employ to estimate the causal impact of these programs on age-four test scores, and Section IV introduces the data and presents an exploratory assessment of whether the design's identifying assumptions hold in the ECLS-B. Section V gives the main impact estimates for universal and targeted pre-K, along with a series of specification checks on these estimates and their difference. Section VI explores potential alternative explanations for the larger impacts of pre-K attendance in states with universal programs beyond universal access per se, and Section VII offers the cost-benefit analysis. Section VIII concludes.

II. Program Landscape

There has been striking growth in public funding of preschool programs since the early 1980s. Figure 1 shows 1968–2015 trends in the number of states funding pre-K programs (left axis) and in enrollment rates of three- and four-year-olds in the federal Head Start program and in any public preschool (right axis).⁶ In the early 1980s, only four states funded pre-K programs; by 2015–2016, this figure reached 43 states and the District of Columbia. Public preschool enrollment rates have risen alongside this increased state funding commitment. This is particularly the case for four-year-olds, for whom enrollment in Head Start—the other primary provider of public preschool—has stagnated since the early 1990s. State-funded pre-K programs indeed focus on four-year-olds: during 2015–2016, 32 percent of four-year-olds were enrolled, compared to 5 percent of three-year-olds (Barnett et al. 2017).⁷

The 2001 birth cohort of the ECLS-B—this study's focus—would have first aged into pre-K eligibility at age four in the fall of 2005, at the start of the 2005–2006 school year.

6. Data on public preschool enrollment rates by age are calculated from the October Current Population Survey School Enrollment supplements (Flood et al. 2015). Head Start enrollment rates divide Head Start enrollments reported by the Head Start Bureau by cohort size estimates based on Census Bureau estimates for July 1, 2005. State funding dates were constructed from program narratives published by NIEER (Barnett et al. 2017).

7. The line between state-funded pre-K and Head Start is sometimes blurred, as some states allow school districts to subcontract with Head Start centers to provide pre-K. The pre-K enrollment measure used in this paper takes into account this possibility, as described in the following.

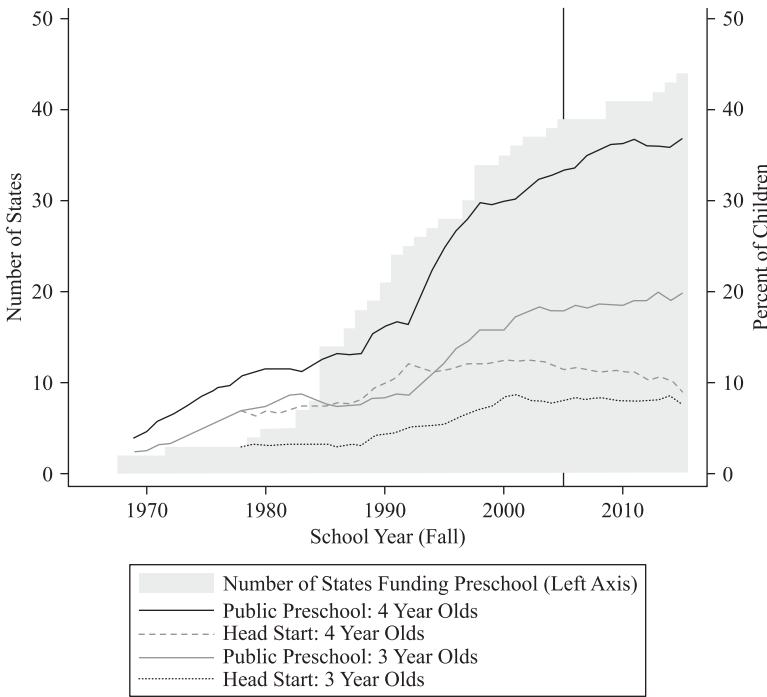
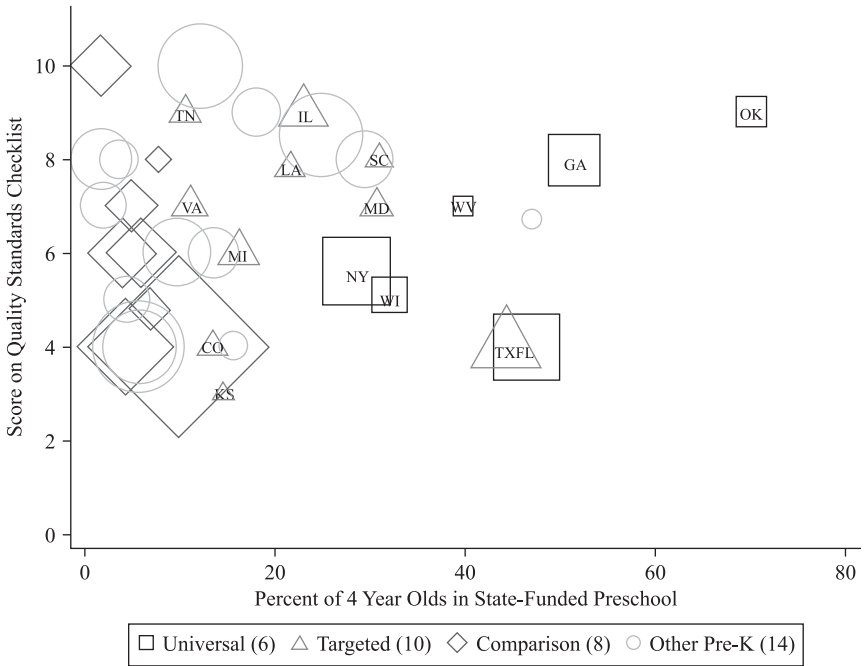


Figure 1
Trends in Public Preschool Enrollment Rates and Pre-K Funding: 1968–2015

Notes: Data on public preschool enrollment rates by age are three-year moving averages calculated from the 1968–2016 October Current Population Survey School Enrollment supplements (Flood et al. 2015). Head Start enrollment rates divide Head Start enrollments reported by the Head Start Bureau by cohort size estimates based on annual (as of July 1) national age-specific population estimates from the Census Bureau. State funding dates were constructed from program narratives published by the National Institute for Early Education Research (Barnett et al. 2017).

At this time (vertical line in Figure 1), state funding for pre-K was not that different than it’s been more recently: 38 states and Washington, DC funded programs. As has also been true in recent years, some programs had no eligibility requirements beyond age (universal programs), whereas others were also means-tested or used other risk factors, like low parental education, to determine eligibility (targeted programs). My main analysis focuses on 16 state pre-K programs where in 2005–2006, there was an enrollment differential favoring four-year-olds of at least eight percentage points and a state-established date by which the youngest enrollees were to have turned age four that did not fall in the middle of a month, according to statistics and program narratives published by the National Institute for Early Education Research (NIEER) (Barnett et al. 2006).⁸

8. I do not consider all states with pre-K programs as treated due to constraints imposed by the ECLS-B and my empirical strategy. See Section III and [Online Appendix A](#).

**Figure 2***Pre-K Access and Quality Standards by State: 2005–2006*

Source: Data are from Barnett et al. (2006).

Notes: Dot sizes represent the size of the state's four-year-old population. The quality standards checklist has ten points, one for each of ten program standards: one point for comprehensive early learning standards, four points for teacher training and credentialing requirements (teacher has BA, specialized training in pre-K, assistant teacher has child development associate degree or equivalent, at least 15 hours of in service training annually), two points for staffing ratios (maximum class size no larger than 20, staff-to-child ratio 1:10 or better), two points for comprehensive services (vision, hearing, health, and one support service, at least one meal provided), and one point for a site visit requirement. To qualify as "treated," a state must have had in 2005–2006 a statewide kindergarten entry cutoff birth date that did not fall in the middle of a month and a pre-K enrollment differential favoring four-year-olds (over three-year-olds) of at least eight percentage points. "Universal" treatment states are ones that met these criteria and had no eligibility requirements beyond age; "targeted" treatment states are ones that met these criteria and had additional eligibility requirements based on family income or other risk factors. (See [Online Appendix Table 1.](#)) Comparison states did not surpass the pre-K enrollment threshold to be a treatment state but did have a statewide kindergarten entry cutoff birth date that did not fall in the middle of a month. Not all comparison states are represented in the figure, as some did not have pre-K programs in 2005–2006. (See [Online Appendix Table 2.](#))

Given these selection criteria, these 16 state pre-K programs were unsurprisingly among the larger ones operating in 2005–2006. In fact, Figure 2 shows that of the five states with the largest pre-K programs in terms of age four enrollment shares at that time—Florida, Georgia, Oklahoma, Texas, and Vermont—only Vermont is excluded from the analysis (due to a having locally determined entry cutoff birth date). The population-weighted average state-funded pre-K enrollment rate for four-year-olds (gap between

age four and age three enrollment rates) across these 16 states was 34.3 percent (32 percentage points) in 2005–2006, compared to only 9.5 percent (5.7 percentage points) in the remaining 22 states with programs.

Figure 2 also denotes (with square markers) which of these 16 states operated universal pre-K in 2005–2006. The six universal states include Georgia and Oklahoma, which have the two longest-standing and most well-studied universal pre-K programs (Gormley and Gayer 2005; Wong et al. 2008; Fitzpatrick 2008, 2010; Cascio and Schanzenbach 2013), whereas the targeted states (triangle markers) include Tennessee, the only state to date with a pre-K program subjected to randomized evaluation (Lipsey et al. 2013; Lipsey, Farran, and Hofer 2016).⁹ Other states represented in Figure 2 had pre-K programs in 2005–2006 but did not meet all of the criteria to be a treatment state. However, eight of these states (diamond markers) meet the criteria to be included in the comparison group, as described in Section III.

While greater political support for universal programs could translate into higher standards or higher per-pupil spending, this does not appear to have been the case at this time. Figure 2 shows no systematic relationship between universality and NIEER's ten-point metric of suggested minimum state pre-K standards, and in fact the targeted programs under study met slightly more of these requirements on average (Online Appendix Table 1). Average per-pupil state spending on pre-K was also about the same in 2005–2006 for the universal and targeted programs under study, at \$3,500–\$3,600 (nominal dollars, population weighted).¹⁰

If not via greater per-pupil resources, how might universal pre-K deliver larger benefits than targeted pre-K, especially for the low-income children that both programs serve? Possible explanations include higher quality teachers¹¹ or higher academic expectations in universal pre-K classrooms,¹² peer effects, or perhaps a different mix of structural inputs for a given level of per-pupil spending. The universal programs under study actually on average imposed less teacher training than targeted ones, possibly in exchange for requiring smaller classes, lower staffing ratios, and more comprehensive services (Online Appendix Table 1).

9. Though universal states tend to have relatively high enrollment rates, the enrollment rate distributions of universal and targeted states overlap. This is expected: universality in this context means only that programs have no means testing in the localities that provide them, not that they are provided in all localities within a state. Classifying states based on enrollment rates rather than mandated eligibility criteria delivers smaller and statistically insignificant group differences in test score impacts, as discussed in Online Appendix A.

10. For most states, this figure represents only state contributions to pre-K. It therefore understates total spending, which could also be funded by local and federal revenue. I account for this possibility in the cost-benefit analysis of Section VII by taking K–12 per-pupil spending as an upper bound. K–12 spending is on average a bit higher in states with universal programs, at \$11,875 per pupil versus \$10,139 per pupil in states with targeted programs.

11. Sabol et al. (2013) find that scores on the Classroom Assessment Scoring System (CLASS) do a better job than inputs (staff qualifications and class size) and learning environment (as measured by the Early Childhood Education Rating System—Revised, or ECERS-R) in predicting test score gains over the pre-K year. Exploiting random assignment of students to kindergarten teachers in Ecuador, Araujo et al. (2016) find that kindergarten teachers with higher CLASS scores have higher value-added for reading and math scores. Araujo, Dormal, and Schady (2019) also show that infants and toddlers quasi-randomly assigned to caregivers with higher CLASS scores in Peru have better fine motor, communication, and problem-solving skills.

12. If prompted to focus on relatively advanced material, teachers may accelerate the learning gains of most students. For example, Engel, Claessens, and Finch (2013) show the more time teachers spend on more advanced mathematics content, the more children gain in math scores over kindergarten year, regardless of demographics.

III. Empirical Strategy

I am interested in the achievement impacts from attending state-funded pre-K programs and the difference in these impacts between universal and targeted programs. At base, the parameters of interest are thus the attendance impacts of each program type. Causal estimates are difficult to obtain. A simple difference in the scores of attendees and nonattendees will be a biased estimate of the attendance effect, since attendance is voluntary, and there might not be state funding to serve all children who are eligible. Children who participate in state-funded pre-K may thus have unobserved characteristics that directly influence their achievement.

I bring an empirical approach to this identification problem that has similar motivations to that taken first in the pre-K evaluation literature by Gormley and Gayer (2005). At base, I wish to compare children who differ in their eligibility for pre-K due to state rules governing age at school entry. Among four-year-olds potentially served by a given pre-K program, those whose fourth birthdays are on or before the birth date threshold to be eligible to attend (for example, September 1) should have a much higher probability of being currently enrolled in pre-K than children whose fourth birthdays are after it. If these two groups of children are on average similar along other dimensions, the difference in their test scores would be the causal effect of being age-eligible for pre-K. Further, scaling this difference by the difference in their pre-K attendance rates would yield the causal effect of pre-K attendance.

One approach to making these two groups of four-year-olds as similar as possible would be to focus on children with birthdays right near the school-entry threshold. However, the ECLS-B precludes such a sharp comparison; even if information on exact birthday were available (it is not), there would be too few observations on a daily basis to generate informative estimates. In past applications using administrative data from specific states or school districts, researchers have addressed this issue by considering a wider range of birth dates around the cutoff, but also recognizing that children with birthdays on opposite sides no longer have the same potential on average. In fact, even if these two groups of children had similar unobservables, they differ along an observed dimension—age—that is strongly related to child development (for example, Elder and Lubotsky 2009). The RD solution is to assume that the test score effects of age or birth date relative to the cutoff are smooth or can be modeled with a polynomial function that is continuous through the cutoff.

The ECLS-B provides information on month of birth, not exact birth date, and data on children across the United States, not just in specific states or school districts. These data support an alternative DD approach—a comparison of the test scores of four-year-olds in adjacent school-entry cohorts in states with the state-funded pre-K programs identified in Section II (the treatment states) versus other states.¹³ I work with 17 such other comparison states (Online Appendix Table 2). The comparison states had a state-established kindergarten entry cutoff birth date in fall 2006 that was not in the middle of the month *and* either had: (i) no state-funded pre-K (nine states) or (ii) state-funded pre-K enrollment rates that were too low or not different enough between three- and four-year-olds,

13. In treatment states, cutoff birthdates for pre-K in fall 2005 are the same as those for kindergarten in fall 2006 (from Barnett et al. 2006), so kids in adjacent pre-K entry cohorts are also in adjacent kindergarten entry cohorts.

according to NIEER, for me to consider them treatment states (the eight states denoted with diamond markers in Figure 2).¹⁴ Differences in the test scores of four-year-olds across adjacent school-entry cohorts in the comparison states are intended to capture what *would* have happened for children in the treatment states in the absence of a state-funded pre-K program, due to aging or other factors.

Ignoring the distinction between universal and targeted programs for now to save on notation, the reduced-form DD model of interest that captures this idea is given by:

$$(1) \quad y_{is} = \theta \text{elig}_{is} \times \text{treat}_s + \sum_{m=-4}^7 \gamma_m \text{elig}_{is}^m + \alpha_s + v_{is}$$

where y_{is} is the age four (2005–2006 academic year in the ECLS-B) test score of child i in state s , and treat_s is a dummy equal to one if s is a treatment state. elig_{is} is then a dummy equal to one if i is in the earlier (or older) entry cohort, set to enter kindergarten in fall 2006 rather than fall 2007 (and pre-K in fall 2005 rather than fall 2006 in treatment states). That is, $\text{elig}_{is} = 1[\text{age}_{i,t} - \text{age}_{k_s}^* \geq 0]$, where $\text{age}_{i,t}$ is child i 's age in months on September 1, 2006, and $\text{age}_{k_s}^*$ is the minimum age in months for kindergarten entry in state s on September 1, 2006.¹⁵ Intuitively, if all states had August 31 or September 1 cutoff birth dates—indeed 22 of the 33 states under study do— elig_{is} would equal one for all ECLS-B respondents born January through August 2001 and zero for those born September through December 2001.

Model 1 is a generalization of the simplest two (group) by two (period) DD model that replaces the direct effects of period (elig_{is}) and group (treat_s) with fixed effects for single months of relative age (the γ_m , where $\text{elig}_{is}^m = 1[\text{age}_{i,t} - \text{age}_{k_s}^* = m]$) and state fixed effects (the α_s), respectively.¹⁶ In principle, I could estimate an alternative triple-difference model that also takes advantage of variation in the timing of school-entry cutoffs across states. Such a model would allow for identification of month-of-birth effects on test scores separately from the effects of academic cohort (Cascio and Lewis 2006)—and separately for treatment and comparison states—through inclusion of month of birth \times treat_s fixed effects. This approach, however, is not possible in this application, since there is so little cross-state variation in the timing of entry cutoffs. That is, month of birth dummies and elig_{is} are highly collinear not just in subgroups of states (for example, among universal states), but also in the full sample.¹⁷ An upside of Model 1, however, is that estimates are

14. For the second group of states, it is impossible to detect a first stage using this empirical approach and the ECLS-B. Note that the 14 states with pre-K programs that are not included in the study (circles in Figure 2) had a cutoff birthdate for school entry that was either locally determined or state determined but in the middle of the month. All comparison states are listed in [Online Appendix Table 2](#), and the selection process is further described in [Online Appendix A](#).

15. Without exact birthday, I must include children who are eligible for kindergarten in fall 2006 (for example, those turning five on September 1, 2006 in a state with a September 1 cutoff date) in the fall 2007 kindergarten cohort. If births are uniformly distributed across days, this should lead to a small attenuation bias in estimates of the eligibility impacts. Thus, I minimize attenuation bias by excluding states with cutoff dates closer to the middle of the month.

16. Substituting a direct effect of elig_{is} for the vector elig_{is}^m in Model 1 yields similar point estimates to those reported later in the paper, but with larger standard errors.

17. As shown in [Online Appendix Tables 1 and 2](#), 22 of 33 states (10 of the 16 treatment states and 12 of the 17 comparison states) require entering prekindergartners to be age four on August 31 or September 1. In addition, five more states (four treatment and one comparison) require prekindergartners to be age four only one month later, on September 30 or October 1. I show below that estimates of Model 1 when limited to states with August 31 or September 1 cutoffs are substantively similar to estimates of Model 1 on the full sample.

not vulnerable to the biases from heterogeneous treatment effects that arise in difference-in-differences settings that exploit variation in treatment timing; see, for example, de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021).

The coefficient of interest in Model 1 is the intent-to-treat (ITT) effect θ , which captures how much more entry cohort, or pre-K eligibility, relates to the age four (2005–2006 academic year) test scores of children in treatment states. But recall that the true parameter of interest is the impact of state-funded pre-K *attendance*, not eligibility. With data from the ECLS-B, I am able to produce estimates of the effects of the treatment on the treated (TOT) by instrumenting for state-funded pre-K attendance, $prek_{is}$, in the model

$$(2) \quad y_{is} = \beta prek_{is} + \sum_{m=-4}^7 \lambda_m elig_{is}^m + \delta_s + \varepsilon_{is}$$

with $elig_{is} \times treat_s$, using two stage least squares (TSLS).

For TSLS estimation of Equation 2 to produce unbiased estimates of β , it must be the case that differences in unobserved determinants of outcomes across adjacent entry cohorts do not systematically differ between the treatment and comparison states, or $elig_{is} \times treat_s$ is uncorrelated with ε_{is} conditional on the $elig_{is}^m$ and state fixed effects. Identification also requires a significant first stage, or that $elig_{is} \times treat_s$ significantly predicts pre-K attendance. The first-stage coefficients on the instrument will by (program) design differ in models that do not distinguish respondents by family income, making TSLS estimation of Model 2 critical to comparing the benefits of universal and targeted programs. I begin by pooling respondents across the family income distribution. However, I quickly move to presenting estimates by family income, since targeted programs should only affect lower-income children.

Being able to estimate TOT impacts is one advantage of using the ECLS-B, as the administrative data used in previous RD applications have been restricted to students who enroll in public pre-K (Lipsey et al. 2015). But there are other advantages to these survey data. For example, the ECLS-B provides a rich set of baseline characteristics, including birth weight and earlier (age two) outcomes, with which to evaluate the internal validity of the empirical approach, and data on alternative care and education options, allowing me to better understand the counterfactual to the state-funded program. The age four (2005–2006) ECLS-B test is also designed to be age appropriate and is administered before children attending pre-K would have progressed to kindergarten, limiting contamination from kindergarten exposure possibly present in past RD studies. The chief limitation of the ECLS-B is small sample size, or statistical power.

IV. Data and Exploratory Analysis

As discussed, the validity of my empirical approach rests on two assumptions. The first is that there is a first-stage relationship between eligibility and state-funded pre-K attendance, or that the interaction of school-entry cohort and residence in a treated state, $elig_{is} \times treat_s$, predicts state-funded pre-K attendance in the 2001 birth cohort as of the 2005–2006 academic year. The second is that this interaction does not predict unobserved correlates of test scores. In this section, I provide some preliminary evidence that these assumptions are met and details on the construction of key variables and the estimation sample from the ECLS-B.

A. State-Funded Pre-K Attendance

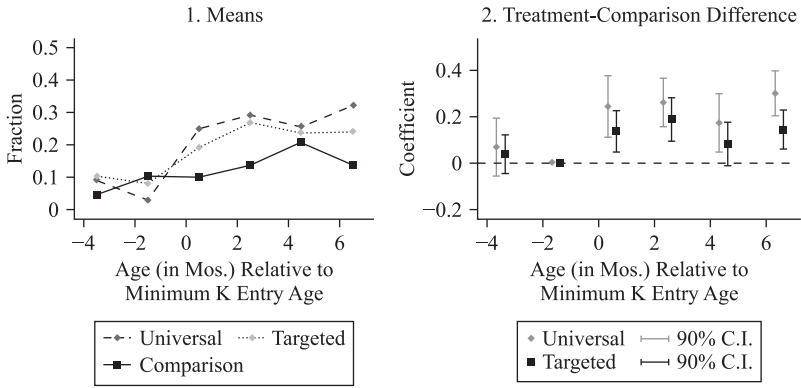
Administrative information on whether the ECLS-B respondents attended state-funded pre-K is unavailable. However, the survey provides detailed information on the care and education of respondents at age four (Wave 3 of the survey, corresponding to 2005–2006) from interviews with both parents and providers. While interviews of providers could in principle offer more reliable information, they were only administered to the one program, center, or person accounting for the *most* of a given child’s nonparental care and thus fail to pick up all cases of enrollment in state-funded pre-K. My base measure of pre-K attendance is therefore the parent report of a child attending free pre-K or free preschool. A limitation of this measure alone, however, is that parents whose children attended a state-funded pre-K program delivered via Head Start may have reported Head Start participation instead. In these cases—in fact in all cases where the provider report signals public pre-K attendance when the parent report does not—I adjust my pre-K attendance variable upward accordingly. I code state-funded pre-K attendance in the provider interviews as public school pre-K or another program (for example, preschool or childcare) sponsored by state or local government or a school district. (See [Online Appendix A](#).)

The first graph in Figure 3, Panel A shows pre-K attendance rates in 2005–2006 by program type in one’s state of residence (universal, targeted, comparison) and age relative to the minimum age for entering kindergarten the subsequent school year (in two-month intervals to reduce noise). Age is increasing along the horizontal axis, with the first two points representing ages of children who would not have been eligible for kindergarten in fall 2006 (or pre-K in fall 2005). The second graph in Figure 3, Panel A then shows the difference in means between each of the two treatment groups and the comparison group, relative to what that difference was for children who just missed eligibility ($-2 \leq \text{age}k_i - \text{age}k_s^* \leq -1$), adjusting for state and month of assessment fixed effects. Aside from age being grouped into intervals, this is a generalization of Model 1, with pre-K attendance as the outcome. For now, the estimation sample includes all children, regardless of family income, who were resident in one of the 16 treatment or 17 comparison states in Wave 3 of the ECLS-B and were five years old between eight months before and four months after their state’s kindergarten entry cutoff.¹⁸

Pooling across family income, the first-stage impacts of age-eligibility should be lower in states with targeted pre-K programs, where eligibility is also based on family income or other risk factors. This expectation is realized in the data. As shown in Panel A, all three groups of states exhibit similar relatively low 2005–2006 pre-K attendance rates among children who were not eligible to attend kindergarten in fall 2006 (or pre-K in 2005 in treatment states). However, pre-K attendance rates between treatment and comparison states diverge for children age-eligible to start pre-K in fall 2005, and the extent of divergence is greater for children in universal programs, as anticipated. As

18. I make this restriction since August 31/September 1 is the modal cutoff date in the sample (see [Online Appendix Tables 1 and 2](#)). I additionally restrict attention to children with nonmissing preschool-age cognitive assessments administered in September or later and nonmissing demographic and background characteristics. These additional restrictions lead to a loss of very few observations. There are 5,100 observations in the sample overall: 1,750 of these children reside in states with targeted programs, 1,150 reside in states with universal programs, and 2,250 reside in the comparison states. Reported sample sizes are rounded to the nearest 50, per IES rules to protect confidentiality of ECLS-B respondents.

Panel A: Pre-K Attendance



Panel B: Preschool-Age Test Scores

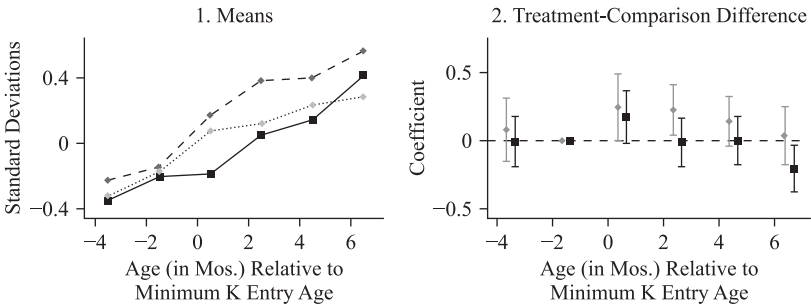


Figure 3

Pre-K Attendance and Test Scores by Age and Program Type

Source: Author’s calculations from the ECLS-B.

Notes: Sample is restricted to respondents with nonmissing values of key variables resident in one of the analysis states at Wave 3 (2005–2006), born within four months after and eight months before that state’s cutoff birth date for kindergarten entry, and assessed during 2005–2006. Panel A corresponds to pre-K attendance in 2005–2006; Panel B corresponds to average standardized reading and math scores in 2005–2006. Subpanel 1 of each panel plots average values of the dependent variable by age relative to the minimum age for kindergarten entry (two-month bins) by state type; see [Online Appendix Tables 1 and 2](#). The dots in Subpanel 2 of each panel represent the coefficients on interactions between a treatment dummy and a series of dummies for age relative to the minimum age for kindergarten entry (two-month bins) from a regression that allows for direct effects of each of these (sets of) variables in addition to month \times year of assessment dummies and state fixed effects. The interaction with the dummy for missing eligibility by one to two months is omitted for identification. Capped vertical lines represent 90 percent confidence intervals, with standard errors clustered on state \times month of birth.

shown in the second graph, the regression-adjusted DD estimates for these age-eligible children are statistically significant.

Columns 1 and 3 in Table 1, Panel A give pre-K attendance rates among age-ineligible children in treatment states (that is, eligible for pre-K in fall 2006, not fall 2005). Columns 2 and 4 show the first-stage DD estimates that correspond to the second subpanel in

Table 1
Descriptive Statistics and Balance Tests on Key Variables, by Program Type: Full Sample

	Universal		Targeted		Universal-Targeted	
	Ineligible Mean [SD] (1)	DD Coeff. (SE) (2)	Ineligible Mean [SD] (3)	DD Coeff. (SE) (4)	DDD Coeff. (SE) (5)	DDD Coeff. (SE) (5)
Panel A: Treatment Variable						
Prekindergarten ^a	0.052	0.211 (0.047)	0.091	0.114 (0.033)		0.097 (0.048)
Panel B: Background Characteristics						
Age in months ^a	48.525 [3.160]	-0.099 (0.060)	48.235 [3.316]	-0.093 (0.051)		-0.006 (0.055)
Female	0.463	0.078 (0.046)	0.521	-0.007 (0.044)		0.085 (0.047)
Black non-Hispanic	0.206	0.042 (0.033)	0.201	0.026 (0.029)		0.015 (0.035)
Hispanic	0.196	-0.034 (0.039)	0.255	-0.021 (0.033)		-0.014 (0.036)
Low birth weight	0.074	0.022 (0.014)	0.079	0.015 (0.012)		0.007 (0.014)
Maternal education \leq HS ^a	0.445	0.022 (0.042)	0.468	-0.031 (0.045)		0.053 (0.042)

(continued)

Table 1 (continued)

	Universal		Targeted		Universal-Targeted	
	Ineligible Mean [SD] (1)	DD Coeff. (SE) (2)	Ineligible Mean [SD] (3)	DD Coeff. (SE) (4)	DDD Coeff. (SE) (5)	DDD Coeff. (SE) (5)
Both biological parents in HH ^a	0.704	-0.052 (0.038)	0.692	0.045 (0.035)	-0.097 (0.041)	-0.097 (0.041)
Not English at home ^a	0.152	-0.010 (0.031)	0.160	-0.008 (0.026)	-0.002 (0.030)	-0.002 (0.030)
Family income ≤ 185% FPL ^a	0.473	-0.092 (0.043)	0.451	-0.064 (0.043)	-0.028 (0.045)	-0.028 (0.045)
Panel C: <i>p</i>-Value: Joint Test						
All background characteristics		0.01		0.17		0.26
Excluding poverty		0.32		0.40		0.27
Observations ^b	300	3,400	400	3,950		7,350

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent-Guardian Interviews,” “Direct Child Assessments,” and “Early Care and Education Providers” (collection at 48 months).

Notes: Odd-numbered columns give means for respondents in treatment states ineligible for pre-K in 2005–2006 [standard deviations for nonbinary variables]. Even-numbered columns give DD coefficients (standard errors) on the interaction between a dummy for being eligible for pre-K in 2005–2006 and a dummy for residing in a treatment state (*elig × treat*) from a separate regression that also includes dummies for state of residence, month × year of assessment, and month age five relative to the state kindergarten entry cutoff/birth date in 2006–2007. See Online Appendix Tables 1 and 2 for lists of universal, targeted, and comparison states. A child is eligible for kindergarten in 2006–2007 if they turned age five in time to start kindergarten in fall 2006, given their date of birth and the kindergarten entry age regulations in effect in 2006–2007 reported by Barnett et al. (2007). Sample is limited to children who turn age five between four months after and eight months before the cutoff date and who were assessed during the 2005–2006 school year. Means and regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state × month of birth.

^aMeasured at preschool age (Wave 3), or in 2005–2006.

^bRounded to the nearest 50, per IES guidelines.

Figure 3, Panel A, that is, subgroup-specific coefficients (standard errors) on $elig_{is} \times treat_s$ from Model 1, which includes fixed effects for state of residence and single months of age relative to the threshold, with $prek_{is}$ as the dependent variable.¹⁹ The estimated gap in 2005–2006 pre-K enrollment rates between the 2006 and 2007 kindergarten cohorts was 21.1 percentage points higher in states with universal pre-K programs than in the comparison states (Column 2). For targeted states, on the other hand, this gap amounted to 11.4 percentage points (Column 4). The difference in these estimates is statistically significant (Column 5).²⁰

B. Demographic and Background Characteristics

Table 1, Panel B gives ineligible means and analogous DD estimates for demographic and background characteristics in the full sample. In addition to basic demographics—age at assessment (in months) and indicators for sex (female) and race (non-Hispanic Black and Hispanic)—I construct indicators for low birth weight (birth weight <2,500 grams), for low maternal education (at or below a high school degree), for a language other than English being spoken in the home, for the presence of both biological parents in the household, and for low family income. I define low income as eligibility for free or reduced-price lunch (family income \leq 185 percent of the federal poverty line [FPL]), since it is the modal eligibility criterion for the targeted states under consideration and thus the best available way to stratify the analysis.²¹

If the TSLS estimates are identified, $elig_{is} \times treat_s$ should have little predictive power with regard to these observed correlates of test scores, just as it should have little predictive power for unobservables. For the most part, the coefficient estimates in the even columns are not statistically significant. There are exceptions, however. Most notably, children who are age-eligible for universal preschool are significantly less likely to be low-income; this leads me to reject the test of joint significance on the DD coefficients (including this variable, the p -value on joint test is 0.01; excluding it, it is 0.32; see Column 2).²² This phenomenon is also evident, but to a lesser extent, among children in targeted states (Column 4). For both subsamples, the DD coefficient for age in months is also negative and (marginally) statistically significant, but small.

These findings, and the fact that some of these coefficients are large even if not statistically significant, suggest the importance of including these background variables as controls in the analysis. They also suggest the importance of seeking additional ways to

19. As in Figure 3, I cluster standard errors on state of residence-by-month of birth (the level of the treatment) and weight the analysis using sampling weights appropriate for analyses using data from the first and third waves of the ECLS-B. The model also includes dummies for month-by-year of the Wave 3 assessment.

20. Reassuringly, the universal-targeted gap in first-stage coefficient estimates ($21.1/11.4=1.85$) is proportionally similar to the universal-targeted gap in the difference in age four and age three state-funded pre-K participation rates reported by NIEER for 2005–2006 ($41.2/23.1=1.78$).

21. The eligibility criterion is relevant for five targeted states—Texas, Maryland, Louisiana, Colorado, and Tennessee—which together account for 56 (69) percent of the four-year-old population (state-funded pre-K enrollment) in the states with targeted programs (Online Appendix Table 1). The remaining states have different income requirements (Michigan, Kansas) or no income requirements, but risk factor requirements that correlate strongly with income (Illinois, South Carolina, Virginia). I test the sensitivity of my conclusions to the definition of poverty below.

22. This is more likely an artifact of sampling variation than of explicit sorting (Dickert-Conlin and Elder 2010).

validate the research design. I do so below both in considering prior (age two) cognitive test scores as an outcome and in testing for “impacts” of eligibility among ineligible students, the idea being that such analyses will only turn up significant coefficients if there is confounding by unobservables. It is important to note, however, that poverty and age are balanced for the *difference* in estimates between universal and targeted states, as demonstrated by the triple-difference (DDD) estimates in Column 5, and it is the difference in program effects that is the focal point of the paper. In addition, among low-income children, who are also the focal group under study, there is balance on observables for universal and targeted programs alike ([Online Appendix Table 3](#)).

V. Effects of Pre-K Eligibility on Preschool-Age Test Scores

A. Baseline Estimates

I focus the analysis of outcomes on cognitive test scores from the third (preschool-age) wave of the ECLS-B. The preschool-age cognitive assessment included math and reading components and was designed to test both for developmental (age-based) milestones and for knowledge and skills considered important for school readiness and early school success (see [Online Appendix A](#)). I standardize test scores to have a mean of zero and a standard deviation of one in the comparison states and calculate average scores across reading and math as an unweighted mean of these standardized scores. Test administration was concentrated during the fall of the 2005–2006 school year, however, possibly raising concerns that not enough time would have elapsed for children’s scores to reflect pre-K. In my robustness analysis, I therefore consider an alternative outcome at preschool age—an indicator for whether a parent reports concern over a child’s readiness for kindergarten. This measure captures parental perceptions of the impacts of pre-K attendance, not just on academic preparation for kindergarten but also behavioral and social preparation as well.²³

Mean preschool-age test scores for the ineligible subsample in treatment states, shown in Column 2 of Table 2, are negative not because universal or targeted states are negatively selected, but rather because there is a strong age gradient in test scores in all states. This is evident in the first graph in Figure 3, Panel B. For comparison states, where the age gradient is least likely to reflect pre-K by design,²⁴ average reading and math test scores rise about 80 percent of a standard deviation between the relatively youngest and relatively oldest children in the estimation sample. There is thus significant variation in early test performance based solely on (relative) age, holding constant family background. This (relative) age gradient in test scores provides a useful benchmark to which I compare the estimated impacts of pre-K attendance.

The remaining curves in Figure 3, Panel B show how the existence of a robust state-funded pre-K program affects this age gradient in preschool-age test scores. There is a clear and relatively sustained divergence between the average test performance of

23. The ECLS-B provides one socioemotional assessment at age four—the two bags task. I analyzed this assessment in an earlier version of this paper, but the estimates were uninformative.

24. The age gradient in comparison states could, however, reflect other characteristics of these states besides their relative lack of state-funded pre-K programs. Fortunately, treatment and comparison states have similar observable demographic and background characteristics on average. See [Online Appendix A](#).

Table 2
Impacts of State-Funded Pre-K on Preschool-Age Test Scores

	Pre-K				
	First-Stage (1)	Ineligible Mean (2)	RF (ITT) (3)	IV (TOT) (4)	OLS (5)
Panel A: Universal (N=3,400)					
No additional controls	0.211 (0.047)	-0.183	0.144 (0.081)	0.682 (0.389)	-0.168 (0.057)
With additional controls ^a	0.211 (0.047)		0.120 (0.065)	0.569 (0.315)	-0.069 (0.048)
Panel B: Targeted (N=3,950)					
No additional controls	0.114 (0.033)	-0.235	-0.009 (0.069)	-0.076 (0.594)	-0.232 (0.049)
With additional controls ^a	0.118 (0.034)		-0.044 (0.060)	-0.373 (0.513)	-0.093 (0.042)
Panel C: p-Value on Universal–Targeted Difference (N=7,350)					
No additional controls	0.045		0.071	0.185	0.227
With additional controls ^a	0.057		0.012	0.046	0.609

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” “Direct Child Assessments,” and “Early Care and Education Providers” (collection at 48 months).

Notes: The first-stage and RF (ITT) [reduced-form (intent-to-treat)] columns in Panels A and B give coefficients on the interaction between a dummy for being eligible for kindergarten in 2006–2007 and a dummy for being in a treated state ($elig \times treat$). (See text or [Online Appendix Tables 1 and 2](#) for definition of universal, targeted, and comparison states.) The IV (TOT) [instrumental variables (treatment-on-treated)] and OLS columns in Panels A and B give coefficients on pre-K attendance ($prek$), in the first case instrumenting with $elig \times treat$ using TSLS. All coefficients are from separate regressions that also include dummies for state of residence, month \times year of assessment, and month age five relative to the state kindergarten entry cutoff birth date in 2006–2007. Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state \times month of birth.

^aAdditional controls include all demographic and background characteristics in Table 1, Panel B: age at assessment, dummies for female, Hispanic, Black non-Hispanic, low birth weight, not English at home, mom has high school degree or less, both biological parents in household, and low-income.

age-eligible children in universal pre-K states relative to the comparison group (Sub-panel 1). The relative gains in test performance are statistically significant for the two youngest groups of children in the eligible cohort (Subpanel 2). While there is some suggestion of such gains for targeted programs, the effect dies out more quickly, and even turns negative for the oldest children in the sample. Combined with the evidence for pre-K attendance in Panel A, these figures are consistent with children gaining more in the short term from pre-K attendance in states with universal programs. Supporting this

inference, the divergence in scores between universal states and the comparison group did not already begin among children who were not yet age-eligible for school (that is, there is no preexisting upward trend).

Table 2 presents the corresponding first-stage and reduced-form (Model 1) estimates of the impacts of pre-K eligibility and IV and OLS (Model 2) estimates of the impacts of pre-K attendance on preschool-age test scores. I show estimates without and then with controls for the demographic and background characteristics in Table 1, Panel B, separately for universal (Panel A) and targeted (Panel B) programs. Though the additional controls have little impact on first-stage estimates, they more moderately impact the reduced-form (RF) and IV estimates. With controls I reject the null that attending a targeted pre-K program has the same effect as attending a universal pre-K program, as shown in Panel C ($p=0.046$). The IV estimated impact of attending pre-K in a universal state also remains substantial, at a marginally statistically significant 0.57 standard deviations—about 70 percent of the impact that the average child would expect from a full year of aging. This estimate is very different from its OLS counterpart (-0.07 standard deviations), which suggests negative selection into pre-K attendance.

B. Heterogeneity by Family Income

I prefer estimates from a sample including children regardless of family income as a starting point, since I cannot perfectly replicate the other eligibility rules for targeted programs in the ECLS-B. Table 3 presents estimates applying arguably the best available consistent definition, which cuts the sample by eligibility for free or reduced-price lunch, based on the preferred specification with the additional controls; group-specific means and balance tests are provided in [Online Appendix Table 3](#). This cut of the data produces income gaps in achievement like those that have been seen in other data (Table 3, Column 2) (Reardon 2011). First-stage impacts of age-eligibility for pre-K (Column 1) are also essentially the same for low-income children regardless of whether a program is universal or targeted (0.226 versus 0.223) but are only evident for children who are not low-income in states with universal programs (0.189 versus 0.037), as expected. [Online Appendix Figure 1](#) provides transparent graphical evidence of these first-stage impacts.

For low-income children, estimated test score gains from pre-K are strikingly different for states with universal versus targeted programs. Indeed, although it remains possible that universal pre-K (Panel A) leads to equivalent gains across the two family-income groups ($p=0.118$ for the RF and $p=0.191$ for IV),²⁵ the estimates for universal programs shown in Table 2 and Figure 3 were clearly driven by the low-income subsample. Age-eligibility for pre-K raises the preschool-age test scores of low-income children in universal states by a significant 0.263 standard deviations (Panel A, Column 3). By contrast, it reduces the test scores of low-income children in targeted states by an insignificant 0.018 standard deviations ($p=0.02$ on the difference; Panel C).²⁶ Figure 4 suggests that this basic insight also holds when I divide the

25. Finding larger effects of pre-K attendance for low-income children is consistent with a framework where higher-income children have relatively high-quality care and education options in the absence of universal pre-K and with much existing evidence on universal preschools both in the United States and worldwide (Cascio 2015; Elango et al. 2016).

26. Estimation of a model that substitutes a “universal” dummy for the *treat_s* in Model 1 and limits the estimation sample to treated states implies that pre-K eligibility yields a relative test score gain (for universal programs) of 0.259 standard deviations (SE=0.009, $p=0.01$). There is thus some precision to be gained from

Table 3
Impacts of State-Funded Pre-K on Preschool-Age Test Scores, by Poverty Status

	Pre-K		Test Score (Average Reading and Math)		
	First-Stage (1)	Ineligible Mean (2)	RF (ITT) (3)	IV (TOT) (4)	OLS (5)
Panel A: Universal					
Low-income (<i>N</i> = 1,550)	0.226 (0.062)	-0.572	0.263 (0.107)	1.160 (0.544)	-0.037 (0.055)
Not low-income (<i>N</i> = 1,850)	0.189 (0.047)	0.165	0.007 (0.102)	0.038 (0.530)	-0.079 (0.077)
<i>p</i> -value on difference	0.550		0.118	0.191	0.635
Panel B: Targeted					
Low-income (<i>N</i> = 1,750)	0.223 (0.051)	-0.497	-0.018 (0.102)	-0.082 (0.450)	-0.015 (0.051)
Not low-income (<i>N</i> = 2,200)	0.037 (0.037)	-0.018	-0.066 (0.080)	no f.s.	-0.174 (0.079)
<i>p</i> -value on difference	0.002		0.732		0.107
Panel C: <i>p</i>-Value on Universal–Targeted Difference					
Low-income	0.996		0.02	0.032	0.728
Not low-income	0.001		0.461	n.a.	0.255

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” “Direct Child Assessments,” and “Early Care and Education Providers” (collection at 48 months).

Notes: A child is considered low-income if their (preschool-age or 2005–2006) family income is at or below 185 percent FPL, the threshold for eligibility for reduced-price lunch and the modal income eligibility criterion for the targeted programs under study. The first-stage and RF (ITT) columns in Panels A and B give coefficients on the interaction between a dummy for being eligible for kindergarten in 2006–2007 and a dummy for being in a treated state (*elig* × *treat*). The IV (TOT) and OLS columns in Panels A and B give coefficients on pre-K attendance (*prek*), in the first case instrumenting with *elig* × *treat* using TSLS. All coefficients are from separate regressions that also include dummies for state of residence, month × year of assessment, and month age five relative to the state kindergarten entry cutoff birth date in 2006–2007, and all demographic and background characteristics listed in Table 1, Panel B except the low-income indicator. Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state × month of birth. “no f.s.” means “no first-stage.” “n.a.” means “not applicable.”

sample by quintiles of a socioeconomic status (SES) index provided by the ECLS-B, derived from factor analysis on parental education, parental occupation, and family income. Test score gains from pre-K in universal states are concentrated in the two

dropping comparison states and focusing on the universal versus targeted difference in effects. However, such an approach precludes estimation of the attendance impacts of each type of program.

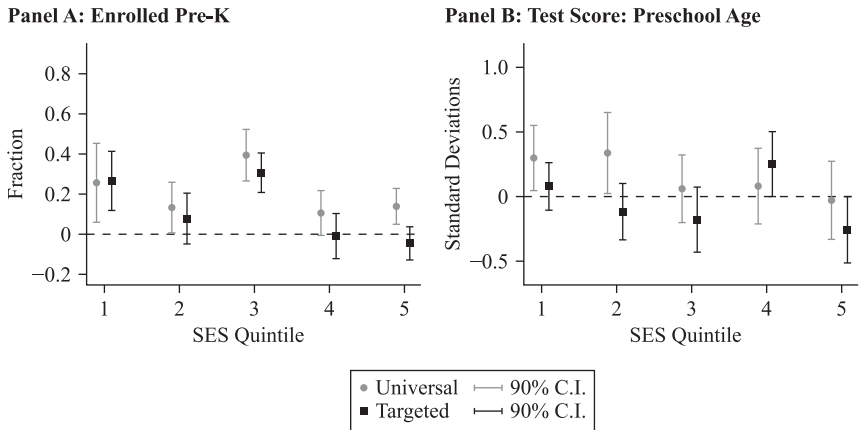


Figure 4

Eligibility Effects on Pre-K Attendance and Test Scores, by Program Type and Socio-economic Status Quintile

Source: Author's calculations from the ECLS-B.

Notes: Sample is restricted to respondents with non-missing values of key variables resident in one of the analysis states at Wave 3 (2005–2006), born within four months after and eight months before that state's cutoff birth date for kindergarten entry, and assessed during 2005–2006. Each dot in each panel represents an estimate of θ in Model 1 restricting attention to children in states with universal or targeted programs (in both cases relative to the same group of comparison states) in the designated quintile of the ECLS-B index for socioeconomic status (SES). The SES index is measured contemporaneously with outcomes and is derived from a factor analysis of parental education, parental occupation, and family income. The underlying regression also includes indicators for month \times year of assessment and the demographic and background characteristics listed in Table 1, Panel B. The capped vertical lines represent 90 percent confidence intervals, with standard errors clustered on state \times month of birth.

bottom quartiles of this index (Panel B) despite positive impacts on pre-K attendance across the income distribution (Panel A).

Returning to Table 3, the IV estimates imply that pre-K attendance raises the preschool-age test scores of low-income children in universal states by 1.16 standard deviations and lowers them by 0.08 standard deviations in targeted states. Previous research suggests that universal programs should have larger effects, but this is a larger difference in attendance effects than we might expect to see based on prior literature. For example, recent estimates suggest that the earliest years of (universal) elementary education raise test performance by up to one standard deviation (for example, Anderson et al. 2011; Fitzpatrick, Grissmer, and Hastedt 2011), much like the IV estimate for universal pre-K.²⁷ However, the average child in the present study has had just a few months of pre-K exposure, not a full school

27. Fitzpatrick, Grissmer, and Hastedt (2011) find that one year of kindergarten or first grade raises reading and math scores by about one standard deviation, using variation in assessment dates in the original kindergarten cohort of the ECLS (ECLS-K). Applying an RD design to exploit age-eligibility rules also in the ECLS-K, Anderson et al. (2011) find that a year of early elementary school exposure raises math scores by 0.75 standard deviations.

year. Likewise, existing estimates for targeted state-funded pre-K programs and Head Start are substantially smaller than those found for elementary education (and universal pre-K). However, though a recent evaluation of the Tennessee Voluntary Pre-K program ultimately delivered negative effects (Lipsey et al. 2013; Lipsey, Farran, and Hofer 2016), these estimates have typically been positive.²⁸

It could be the case that the existing evaluation literature understates the expected difference in attendance impacts between universal and targeted pre-K programs—and not just due to differences in methodology, outcomes, timing, and counterfactual enrollment patterns across studies. For example, while the attendance impacts of universal pre-K seem quite large given just a few months of exposure, other research suggests that learning may decelerate over the school year (Kuhfeld and Soland 2021) and points out that the same intervention typically yields larger test score gains at younger ages (Cascio and Staiger 2012). The Tennessee program also scores higher on NIEER's quality checklist than the average targeted state in the present study ([Online Appendix Table 1](#)). That said, the limited statistical power afforded by the ECLS-B means that I am more confident in there being a difference in impacts between universal and targeted programs—and in the impacts of universal programs in particular being positive—than I am in their magnitudes.

Figure 5 presents graphs of the test score impacts of pre-K by family income in an analogous way to Figure 3, Panel B. The concentration of positive impacts of universal pre-K among the youngest eligible students is now even more evident (Panel A). Separate analyses of the subcomponents of the test ([Online Appendix Figures 2 and 3](#)) reveal that performance on the math subcomponent is responsible for this pattern; effects on reading scores are more sustained. Moreover, estimating the preferred model separately for the standardized reading and math scores, as done in Panels B and C of Table 4 for the full sample ([Online Appendix Table 4](#) for the low-income subsample), reveals universal pre-K effects that are much more precisely estimated for reading scores. Preexisting trends in reading scores are also relatively similar across the three groups of states ([Online Appendix Figure 2](#)), lending greater credibility to those findings.

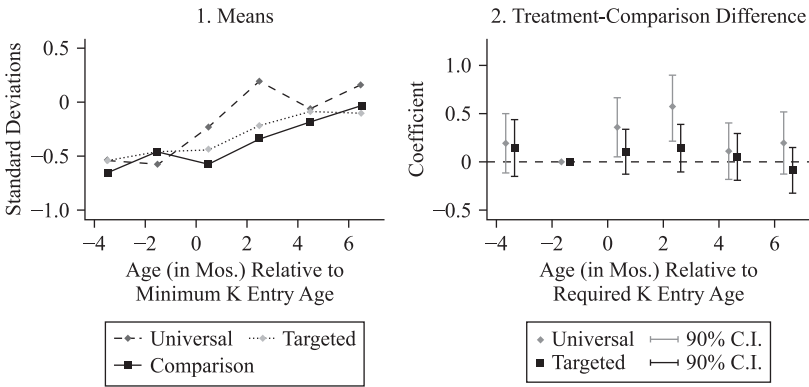
C. Specification Checks

The estimates presented thus far suggest that children attending pre-K in states with universal programs experience larger early test score gains than children attending pre-K in states with targeted programs, and the difference is more pronounced for the low-income children who would likely meet eligibility criteria for either type of program. But these findings could still be an artifact of the research design, estimation sample, or choice of outcome. In this section, I assess the robustness of the basic set of results to these decisions.

In Table 4, Panel D, I first consider impacts of pre-K attendance on a test of mental development at age two (or in Wave 2 of the ECLS-B), before children would have been

28. For example, using randomized variation in Head Start attendance from the Head Start Impact Study, Kline and Walters (2016) estimate that one year of Head Start attendance raises the average of standardized reading and math scores by around 0.25 standard deviations. An evaluation of the Tennessee program also found short-term effects of attendance around a third of a standard deviation (Lipsey, Farran, and Hofer 2016).

Panel A: Low-Income ($\leq 185\%$ FPL)



Panel B: Not Low-Income ($>185\%$ FPL)

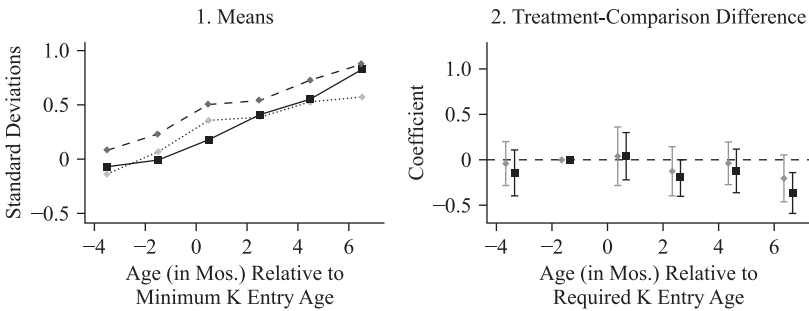


Figure 5

Pre-K Eligibility and Test Scores by Age, Program Type, and Poverty Status

Source: Data are from the ECLS-B.

Notes: Sample is restricted to respondents with nonmissing values of key variables resident in one of the analysis states at Wave 3 (2005–2006), born within four months after and eight months before that state’s cutoff birth date for kindergarten entry, and assessed during 2005–2006. The dependent variable in each panel is average standardized reading and math scores during Wave 3, when respondents were four years of age. Panel A corresponds to respondents who were eligible for free or reduced-price lunch in 2005–2006; Panel B corresponds to respondents who were not. Subpanel 1 of each panel plots the average standardized test score by age relative to the minimum age for kindergarten entry (two-month bins) by state type; see [Online Appendix Tables 1 and 2](#). The dots in Subpanel 2 of each panel represent the coefficients on interactions between a treatment dummy and a series of dummies for age relative to the minimum age for kindergarten entry (two-month bins) from a regression that allows for direct effects of each of these (sets of) variables in addition to month \times year of assessment dummies and state fixed effects. The interaction with the dummy for missing eligibility by one to two months is omitted for identification. Capped vertical lines represent 90 percent confidence intervals, with standard errors clustered on state \times month of birth.

Table 4
Sensitivity of Estimated Effects of Pre-K to the Choice of Outcome

	Pre-K	Test Scores		
	First-Stage (1)	RF (ITT) (2)	IV (TOT) (3)	OLS (4)
Panel A: Baseline				
Universal (<i>N</i> = 3,400)	0.211 (0.047)	0.120 (0.065)	0.569 (0.315)	-0.069 (0.048)
Targeted (<i>N</i> = 3,950)	0.118 (0.034)	-0.044 (0.060)	-0.373 (0.513)	-0.093 (0.042)
<i>p</i> -value on difference	0.057	0.012	0.046	0.609
Panel B: Reading Score Only				
Universal (<i>N</i> = 3,400)	0.211 (0.047)	0.135 (0.064)	0.639 (0.328)	-0.036 (0.047)
Targeted (<i>N</i> = 3,950)	0.118 (0.034)	-0.037 (0.067)	-0.314 (0.564)	-0.100 (0.045)
<i>p</i> -value on difference	0.057	0.007	0.054	0.214
Panel C: Math Score Only				
Universal (<i>N</i> = 3,400)	0.211 (0.047)	0.106 (0.082)	0.500 (0.378)	-0.102 (0.058)
Targeted (<i>N</i> = 3,950)	0.118 (0.034)	-0.051 (0.067)	-0.432 (0.587)	-0.087 (0.047)
<i>p</i> -value on difference	0.057	0.061	0.102	0.786
Panel D: Mental Score (age 2)^a				
Universal (<i>N</i> = 3,350)	0.211 (0.046)	0.005 (0.069)	0.023 (0.326)	-0.062 (0.051)
Targeted (<i>N</i> = 3,950)	0.115 (0.034)	0.035 (0.068)	0.299 (0.610)	-0.081 (0.051)
<i>p</i> -value on difference	0.049	0.613	0.556	0.643

(continued)

Table 4 (continued)

	Pre-K	Test Scores		
	First-Stage (1)	RF (ITT) (2)	IV (TOT) (3)	OLS (4)
Panel E: = 1 If Parent Reports Not Ready for K (Age 4)				
Universal ($N=3,400$)	0.207 (0.047)	-0.091 (0.046)	-0.440 (0.231)	0.003 (0.025)
Targeted ($N=3,950$)	0.114 (0.033)	-0.036 (0.036)	-0.315 (0.343)	-0.021 (0.023)
<i>p</i> -value on difference	0.053	0.261	0.712	0.327

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” “Direct Child Assessments,” and “Early Care and Education Providers” (collection at 48 months).

Notes: The first-stage and RF (ITT) columns give coefficients on the interaction between a dummy for being eligible for kindergarten in 2006–2007 and a dummy for being in a treatment state ($elig \times treat$). The IV (TOT) and OLS columns give coefficients on pre-K attendance ($prek$), in the first case instrumenting with $elig \times treat$ using TSLS. All coefficients are from separate regressions that also include dummies for state of residence, month \times year of assessment, and month age five relative to the state kindergarten entry cutoff birth date in 2006–2007, and all demographic and background characteristics listed in Table 1, Panel B. Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state \times month of birth.^a Additional controls include age at assessment and dummies for month \times year of assessment in Wave 2 of the ECLS-B, and panel weights incorporate observation in Wave 2.

eligible for pre-K.²⁹ Performance on this test is neither significantly affected by pre-K attendance nor significantly different across states with universal and targeted programs ($p=0.613$), suggesting limited contamination by unobservables. The estimates for age-four test impacts are also similar when age-two scores are included as controls ([Online Appendix Table 5](#)).³⁰

In Table 4, Panel E, I then consider an alternative outcome at age four—an indicator for whether a parent reports that their child is not ready for kindergarten. Mirroring the estimates for test scores, the parents of children attending pre-K in states with universal programs are less likely to express concern over their kindergarten readiness. The estimates are more pronounced among lower income parents ([Online Appendix Table 4](#)), but as was also the case with test scores, I cannot rule out that they are identical across

29. The Wave 2 cognitive assessment is the Bayley Short Form–Research Edition (based on the Bayley Scales of Infant Development 2nd Edition). Results with Bayley motor scores as an outcome show a similar pattern as those for mental scores and so are omitted for brevity. The model in Panel D also includes age of assessment and dummies for month-by-year of assessment in the relevant wave, in addition to the controls of the baseline model, and is weighted by sampling weights appropriate for inclusion of that wave.

30. Controlling for age-two scores in the full sample has no appreciable effect on the estimates for universal programs but raises estimates for targeted programs to an extent that the gap in IV estimates across program types is no longer statistically significant ($p=0.178$). The difference in IV estimates for low-income children does, however, remain marginally significant with these controls included ($p=0.055$).

Table 5
Sensitivity of Estimated Effects of Pre-K on Test Scores to Estimation Sample

	Pre-K	Test Scores		
	First-Stage (1)	RF (ITT) (2)	IV (TOT) (3)	OLS (4)
Panel A: Baseline				
Universal (<i>N</i> = 3,400)	0.211 (0.047)	0.120 (0.065)	0.569 (0.315)	-0.069 (0.048)
Targeted (<i>N</i> = 3,950)	0.118 (0.034)	-0.044 (0.060)	-0.373 (0.513)	-0.093 (0.042)
<i>p</i> -value on difference	0.057	0.012	0.046	0.609
Panel B: Sample ±4 Months from Threshold^a				
Universal (<i>N</i> = 2,000)	0.226 (0.054)	0.203 (0.075)	0.900 (0.350)	-0.050 (0.071)
Targeted (<i>N</i> = 2,350)	0.129 (0.039)	0.042 (0.068)	0.321 (0.527)	-0.067 (0.056)
<i>p</i> -value on difference	0.091	0.024	0.197	0.808
Panel C: Identification ±4 Months from Threshold^b				
Universal (<i>N</i> = 3,400)	0.218 (0.055)	0.220 (0.074)	1.007 (0.377)	-0.066 (0.047)
Targeted (<i>N</i> = 3,950)	0.143 (0.039)	0.040 (0.070)	0.278 (0.488)	-0.092 (0.042)
<i>p</i> -value on difference	0.191	0.015	0.104	0.587
Panel D: Expanded Sample (+ Middle-of-Month Cutoffs)^c				
Universal (<i>N</i> = 3,450)	0.212 (0.045)	0.127 (0.065)	0.600 (0.314)	-0.046 (0.048)
Targeted (<i>N</i> = 4,250)	0.118 (0.032)	-0.020 (0.059)	-0.170 (0.503)	-0.089 (0.040)
<i>p</i> -value on difference	0.042	0.022	0.089	0.379

(continued)

Table 5 (continued)

	Pre-K	Test Scores		
	First-Stage (1)	RF (ITT) (2)	IV (TOT) (3)	OLS (4)
Panel E: Restricted Sample (Aug. 31/Sept. 1 Cutoffs)^d				
Universal ($N=1,800$)	0.267 (0.054)	0.150 (0.085)	0.561 (0.311)	-0.086 (0.072)
Targeted ($N=1,850$)	0.192 (0.037)	-0.038 (0.076)	-0.201 (0.397)	-0.095 (0.061)
<i>p</i> -value on difference	0.173	0.010	0.023	0.908

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” “Direct Child Assessments,” and “Early Care and Education Providers” (collection at 48 months).

Notes: See note to Table 4.

^aSample further limited to respondents with birthdays within four months of the cutoff birth date for kindergarten entry in their Wave 3 state of residence.

^bInteractions between *treat* and indicators for birth four to five months and six to seven months after cutoff included.

^cSample expanded to include respondents residing in treatment states with middle-of-month cutoffs (targeted: AR, NC; universal: ME) not born in the cutoff birth date month.

^dSample limited to treatment and comparison states with Aug. 31 or Sept. 1 cutoffs ([Online Appendix Tables 1 and 2](#)).

income groups ($p=0.35$). Unlike in the test score case, however, I also cannot rule out that the IV estimates are the same across program types for low-income children ($p=0.22$), though the evidence is slightly more suggestive for a reduction in concern over *academic* readiness for kindergarten in particular ($p=0.17$). Similar findings do not arise for parent reports of concern over school readiness on other margins ([Online Appendix Table 6](#)), suggesting that the relative benefits of universal pre-K may be primarily academic. But the relative lack of statistical power makes all of these complementary estimates suggestive at best.

I consider several changes to the estimation sample in Table 5, returning to the original test score outcome. I considered a wider age span for eligible children in the baseline estimation sample (eight months) than for ineligible children (four months) to improve statistical power, but possibly at the expense of greater bias. Limiting estimation to children with birthdays within four months of the age-eligibility threshold (Panel B), who are more similar to one another along some but not all dimensions ([Online Appendix Table 7](#)), actually generates more positive estimates for the impacts of pre-K eligibility and attendance, enough so that I am less confident that the IV estimates differ between universal and targeted states in the full sample ($p=0.197$). However, I remain reasonably confident in this conclusion for low-income children, even in this restricted sample ($p=0.060$; [Online Appendix Table 8](#)). In Panel C, I approach this concern in a different way, maintaining the full sample, but controlling directly for interactions

between $treat_{is}$ and $elig_{is}^{4-5}$ and $elig_{is}^{6-7}$ (that is, indicators for the oldest relative age groups in the sample). This alternative model only exploits variation in eligibility within four months of the threshold but maintains the full sample to identify coefficients on the controls. The findings are similar.

The two remaining changes to the estimation sample represented in Table 5 make it more, then less, expansive. Adding treatment states may further help limit the influence of idiosyncratic state samples, since the ECLS-B is not designed to be state representative. With this in mind, I expand the sample to include children in the two targeted states (Arkansas and North Carolina) and one universal state (Maine) with middle of the month birth date cutoffs, as long as their birth date is not in the cutoff month (to minimize misclassification). Results are not much changed (Panel D). Further, limiting attention to the ten treatment states (five universal, five targeted) and 12 comparison states with cutoff birth dates on August 31 or September 1—thus focusing on the largest subsample where eligibility and month of birth are collinear—barely changes the IV estimates in the full sample (Panel E). Moreover, in both the full sample and the low-income subsample (Online Appendix Table 8), state-funded pre-K attendance in universal states continues to yield significantly larger test score gains.

VI. Interpretation

It may be tempting to conclude that universal access is itself the driver of the generally robust finding of larger effects of pre-K attendance in states with universal programs. However, whether a pre-K program is universal or not is not randomly assigned; state contexts, pre-K programs, and populations differ along other dimensions. In this section, I attempt to rule out leading alternative explanations—differences in counterfactual care, other pre-K characteristics (specifically maximum class size), and demographics in states with universal programs. I also present additional evidence consistent with an access interpretation: The estimates for universal pre-K look much like what one finds for universal kindergarten, at age five, within the ECLS-B.

A. Differences in the Counterfactual?

Existing literature suggests that if universal pre-K attendees were drawn less from other center-based care and more from informal or parental care, we would see relatively large test score effects of universal pre-K, all else constant.³¹ I estimate the RF impact of age-eligibility for pre-K and the IV effect of pre-K attendance for a mutually exclusive set of alternative care options—Head Start, other center-based care, informal nonparental care, and parental care.³² For additional statistical power, I also combine Head Start and other center-based care as “formal care” and informal nonparental and parental care as “informal care.”

31. In analyses of data from the Head Start Impact Study, Feller et al. (2016) and Kline and Walters (2016) find that Head Start has much smaller impacts on children who would have otherwise been in center-based care.

32. The reduced-form DD coefficients therefore add up to zero across all categories, including the first-stage DD coefficient for pre-K. Reassuringly, reported Head Start enrollment rates for low-income children in universal and targeted states (at 12 percent) are quite similar to what they are in the administrative data.

The IV estimates for the full sample (Online Appendix Table 9) imply that approximately 90 percent of targeted pre-K attendees were drawn from informal care versus 46 percent of universal pre-K attendees. The difference is not surprising. Universal programs serve higher income children, for whom formal care or education in the absence of state-funded pre-K is relatively common. More telling are findings for the low-income subsample, where the difference in estimated attendance effects across program types is most pronounced. Table 6 shows substitution patterns that are more similar across program types in this sample, with point estimates implying that about 35 percent of low-income pre-K attendees would have otherwise attended Head Start, and about 15 percent would have otherwise been in informal, nonparental care. But parental care was a more likely alternative in the targeted case, so that 70 percent of low-income targeted pre-K attendees would have otherwise been in informal care, compared to 55 percent of low-income universal pre-K attendees.³³ Though the difference is not statistically significant ($p=0.66$), this suggests the estimated gap in pre-K attendance effects across program types is actually *lower* than it would be with a constant counterfactual. It therefore does not provide strong evidence against the conclusion that universal programs outperform targeted ones.

B. Differences in Other Characteristics of Programs or Populations?

As noted in Section II, though universal and targeted programs look similar in the aggregate in terms of available resources and the overall number of minimum standards to which they are held, the two types of programs differ in how funds are allocated or which standards are emphasized on average, and it might be these differences in resource allocation that are driving the differential in estimated effects across states with different pre-K access. Of particular interest is that the universal programs under study prioritize smaller class sizes, which in early education have been convincingly shown to produce higher immediate test scores (Krueger 1999). With the caveat that the decision to require small class sizes could be endogenous to the access decision itself, I can investigate whether the difference in RF and IV estimates across universal and targeted states is robust to allowing for heterogeneity in pre-K impacts by this program dimension. Using a similar approach, I can also explore robustness to regression adjusting for heterogeneity in pre-K impacts by the demographic and background characteristics in Table 1, Panel B.³⁴

33. These figures for program substitution are similar to those found for the recent Head Start Impact Study (Feller et al. 2016; Kline and Walters 2016).

34. Specifically, I begin with the triple-difference RF model:

$$y_{isp} = \theta_{UT} \text{elig}_{is} \times \text{treat}_s \times \text{uni}_s + \theta_T \text{elig}_{is} \times \text{treat}_s + \theta_D \text{elig}_{is} \times \text{treat}_s \times C_s + \delta \text{elig}_{is} \times C_s + \sum_{p=0}^1 \sum_{m=-4}^7 \gamma_{mp} \text{elig}_{is}^m \times P_p + \alpha_{sp} + \sum_{p=0}^1 (x_{is} \times P_p)' \pi_p + \omega_{isp}$$

where uni_s and treat_s represent, respectively, a dummy for whether s is a universal (treated) state and a dummy for whether s is a treatment state at all (universal or targeted), and C_s is the number of class size standards (out of two possible) that the state program requires. All of the controls are interacted with an indicator for the estimation sample, $P_j = 1 [j = p]$, with $p = 1$ for the universal pre-K estimation sample and $p = 0$ for targeted pre-K estimation sample. Excluding the interactions with the class size variable ($\text{elig}_{is} \times \text{treat}_s \times C_s$ and $\text{elig}_{is} \times C_s$), θ_{UT} is then the difference in the reduced-form DD estimates for universal and targeted pre-K programs presented in Table 2, and

Table 6
Impacts on Alternative Care Arrangements, by State Program Type: Low-Income Children

	Pre-K						
	First-Stage (1)	Head Start (2)	Other Center-Based Care (3)	Any Formal (2+3) (4)	Informal Nonparental Care (5)	Parental Care (6)	Any Informal (5+6) (7)
Panel A: Ineligible Means							
Universal	0.08	0.22	0.21	0.43	0.18	0.31	0.49
Targeted	0.11	0.21	0.10	0.32	0.18	0.40	0.57
Panel B: Reduced-Form, Coeff. (SE) on $elig \times treat_s$							
Universal ($N = 1,550$)	0.226 (0.062)	-0.079 (0.070)	-0.022 (0.058)	-0.102 (0.088)	-0.033 (0.047)	-0.092 (0.072)	-0.125 (0.068)
Targeted ($N = 1,750$)	0.223 (0.051)	-0.082 (0.063)	0.016 (0.051)	-0.066 (0.069)	-0.038 (0.045)	-0.120 (0.059)	-0.158 (0.060)
<i>p</i> -value on diff.	0.966	0.968	0.483	0.711	0.924	0.700	0.647

(continued)

Table 6 (continued)

	Pre-K		Alternatives				
	First-Stage (1)	Head Start (2)	Other Center-Based Care (3)	Any Formal (2+3) (4)	Informal Nonparental Care (5)	Parental Care (6)	Any Informal (5+6) (7)
Panel C: Instrumental Variables, Coeff. (SE) on $prek_{it}$							
Universal ($N = 1,550$)	n.a.	-0.349 (0.274)	-0.099 (0.242)	-0.448 (0.317)	-0.146 (0.203)	-0.406 (0.327)	-0.552 (0.317)
Targeted ($N = 1,750$)	n.a.	-0.368 (0.244)	0.073 (0.227)	-0.295 (0.273)	-0.168 (0.196)	-0.537 (0.268)	-0.705 (0.273)
p -value on diff.	n.a.	0.948	0.459	0.660	0.914	0.704	0.660

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, "Children's Birth Certificates" (collection at nine months), "Parent-Guardian Interviews," "Direct Child Assessments," and "Early Care and Education Providers" (collection at 48 months).

Notes: A child is considered low-income if their (preschool-age or 2005-2006) family income is at or below 185 percent FPL, the threshold for eligibility for reduced-price lunch and the modal income eligibility criterion for the targeted programs under study. RF coefficients in Panel B are on the interaction between a dummy for being eligible for kindergarten in 2006-2007 and a dummy for being in a treated state ($elig \times treat$). IV coefficients in Panel C are on pre-K attendance ($prek$), estimated by instrumenting with $elig \times treat$ using TSLS. All coefficients are from separate regressions that also include dummies for state of residence, month \times year of assessment, and month age five relative to the state kindergarten entry cutoff birth date in 2006-2007, and all demographic and background characteristics listed in Table 1, Panel B except the low-income indicator. Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state \times month of birth. "n.a." means "not applicable."

Table 7, Panel A presents the difference in RF (Columns 1 and 4) and IV (Columns 2 and 5) estimates across universal and targeted programs from the model with additional controls, along with estimates of the coefficients capturing how class size standards influence the effect of pre-K eligibility (Columns 3 and 6), both for the full sample and for the low-income subsample.³⁵ The universal versus targeted difference in RF estimates shrinks somewhat, as pre-K eligibility has a greater impact on test scores when class sizes are required to be small, as anticipated. But the difference in IV estimates remains at least as large in both samples as it was at baseline. The estimates in Table 7, Panel B similarly show demographic heterogeneity in the impacts of pre-K exposure in the expected direction. Effects of pre-K eligibility tend to be larger (and are often statistically significant) for more disadvantaged populations, like children of mothers with no more than a high school degree. However, the populations of universal and targeted states are similar enough that this has no large impact on the conclusion that universal programs outperform targeted ones.

C. Comparison to the Impacts of Kindergarten Using an Alternative Approach

To provide further evidence consistent with an access interpretation, I examine whether the test score impacts from attending universal public kindergarten look substantively similar to those from attending universal pre-K in the ECLS-B.³⁶ The fourth (kindergarten-age) wave of the ECLS-B also included math and reading cognitive assessments, again largely administered in the fall of the academic year, which I standardized and aggregated in a similar fashion to the preschool assessments. It also included information on grade of enrollment and whether the school was public or private (if applicable).

To carry out this analysis, however, I must rely on a different empirical approach than that earlier in the paper. In particular, since universal kindergarten is available in public schools across all states, I can only exploit age-eligibility rules using an RD model in the RF, given by:

$$(3) \quad y_{is} = \tilde{\theta}_{elig_{is}} + f(\text{relage}_{is}; \tilde{\lambda}) + \tilde{\alpha}_s + \varpi_{is}$$

$f(\text{relage}_{is}; \tilde{\lambda})$ represents some smooth function, with parameter vector $\tilde{\lambda}$, of the difference between child i 's age in months on September 1, 2006 and the minimum required age in state s on that date for kindergarten entry; that is, $\text{relage}_{is} = \text{age}_{ki} - \text{age}_{ks}^*$. For simplicity and because it is a good fit to the data (see [Online Appendix Figure 4](#)), I specify the smooth function in age relative to the cutoff as linear within plus or minus four months of the entry age threshold, with a different slope among those eligible for

θ_T is the baseline reduced-form DD estimate for targeted programs. Of interest is then estimation with the class size interactions included, which adjusts estimates of θ_{UT} for the correlation between a state program offering universal access and the program requiring small classes. The coefficients on pre_{is} and $\text{pre}_{is} \times \text{uni}_s$ in the structural model of interest are estimated using TSLS with $\text{elig}_{is} \times \text{treat}_s \times \text{uni}_s$ and $\text{elig}_{is} \times \text{treat}_s$ as excluded instruments. The model with demographic heterogeneity replaces C_s with X_i , where X_i is one of the characteristics in Table 1, Panel B. I also estimate version of this model including multiple characteristics simultaneously.

35. With reference to the reduced-form model presented in the prior footnote, Columns 1 and 4 present estimates of θ_{UT} , and Columns 3 and 6 present estimates of θ_D .

36. As noted, the magnitude of the impacts of universal pre-K attendance in the preferred model appears comparable to that of universal early education more generally, but the relevant studies (Anderson et al. 2011; Fitzpatrick, Grissmer, and Hastedt 2011) used a different data source, the ECLS-K.

Table 7
Robustness of the Universal-Targeted Difference in Age Four Test Estimates to Adjusting for Other Sources of State Heterogeneity

	Full Sample			Low-Income		
	Test Scores		Characteristic	Test Scores		Characteristic
	RF (ITT) (1)	IV (TOT) (2)	RF (3)	RF (ITT) (4)	IV (TOT) (5)	RF (6)
Baseline	0.164 (0.065)	0.942 (0.473)	n.a.	0.281 (0.120)	1.241 (0.580)	n.a.
Panel A: Heterogeneity by Class Size Requirements						
Class size standards	0.135 (0.071)	1.154 (0.521)	0.125 (0.067)	0.231 (0.114)	1.196 (0.590)	0.214 (0.108)
Panel B: Heterogeneity by Individual Background Characteristics						
Female	0.162 (0.065)	1.090 (0.498)	0.088 (0.063)	0.279 (0.120)	1.235 (0.576)	0.062 (0.095)
Black non-Hispanic	0.163 (0.065)	0.994 (0.485)	0.121 (0.091)	0.282 (0.121)	1.248 (0.576)	0.200 (0.114)
Hispanic	0.154 (0.066)	0.907 (0.530)	0.023 (0.086)	0.267 (0.120)	1.190 (0.600)	-0.058 (0.113)
Low birth weight	0.164 (0.065)	0.952 (0.479)	0.022 (0.083)	0.281 (0.120)	1.243 (0.578)	0.102 (0.097)

(continued)

Table 7 (continued)

	Full Sample			Low-Income		
	Test Scores		Characteristic	Test Scores		Characteristic
	RF (ITT) (1)	IV (TOT) (2)	RF (3)	RF (ITT) (4)	IV (TOT) (5)	RF (6)
Maternal education \leq HS	0.160 (0.065)	1.313 (0.662)	0.172 (0.074)	0.277 (0.121)	1.233 (0.567)	0.200 (0.103)
Both biological parents in HH	0.157 (0.065)	0.674 (0.449)	-0.143 (0.072)	0.276 (0.119)	1.228 (0.579)	-0.059 (0.096)
Not English at home	0.165 (0.065)	1.010 (0.493)	0.119 (0.089)	0.282 (0.120)	1.255 (0.577)	0.111 (0.114)
Family income \leq 185% FPL	0.168 (0.067)	1.314 (0.691)	0.116 (0.072)	n.a.	n.a.	n.a.
All characteristics	0.143 (0.066)	1.104 (0.703)	n.s.	0.256 (0.119)	1.198 (0.598)	n.s.

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, "Children's Birth Certificates" (collection at nine months), "Parent-Guardian Interviews," "Direct Child Assessments," and "Early Care and Education Providers" (collection at 48 months).

Notes: A child is considered low-income if their (preschool-age or 2005-2006) family income is at or below 185 percent FPL. RF (ITT) columns give the difference in RF (ITT) coefficients on $elig \times treat$ between universal and targeted states. IV (TOT) columns give the difference in IV (TOT) coefficients on $prek$ (instrumented with $elig \times treat$ using TSLS) between universal and targeted states. The RF column shows the coefficient on $elig \times treat \times characteristic$ from the same model that produced the RF (ITT) coefficients. All regressions include dummies for state of residence, month \times year of assessment, and month age five relative to the state kindergarten entry cutoff birth date in 2006-2007, and all demographic and background characteristics listed in Table 1, Panel B (except the low-income indicator in Columns 4-6), with separate control coefficients for the universal and targeted estimation samples. Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state \times month of birth. "n.a." means "not applicable." "n.s." means "not shown."

Table 8
*Impacts of Universal Public Kindergarten on Kindergarten Age Test Scores,
 Overall and by Poverty Status*

	Public K	Test Score (Average Reading and Math)		
	First-Stage (1)	RF (ITT) (2)	IV (TOT) (3)	OLS (4)
Panel A: Full Sample ($N = 2,400$)				
All states	0.657 (0.036)	0.376 (0.084)	0.573 (0.132)	0.250 (0.064)
Panel B: Low-Income ($N = 1,050$)				
All states	0.778 (0.050)	0.440 (0.144)	0.566 (0.182)	0.458 (0.107)
Panel C: Not Low-Income ($N = 1,300$)				
All states	0.557 (0.049)	0.332 (0.095)	0.597 (0.180)	0.110 (0.073)

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” (collection at 48 months), “Direct Child Assessments,” “Early Care and Education Providers,” and “School Questionnaires” (collection at kindergarten entry).

Notes: A child is considered low-income if their (preschool-age or 2005–2006) family income is at or below 185 percent FPL, the threshold for eligibility for reduced-price lunch. The first-stage and RF (ITT) columns in Panels A and B give coefficients on a dummy for being eligible for kindergarten in 2006–2007 (*elig*). The IV (TOT) and OLS columns in Panels A and B give coefficients on public kindergarten attendance, in the first case instrumenting with *elig* using TSLS. All coefficients are from separate regressions that also include as controls dummies for state of (Wave 4) residence and month \times year of (Wave 4) assessment, a linear term in month age five relative to the state kindergarten entry cutoff birth date in 2006–2007, an interaction between that linear term and *elig*, and all demographic and background characteristics listed in Table 1, Panel B (except the low-income dummy in Panels B and C). Regressions are weighted by longitudinal sampling weights, and standard errors (in parentheses) are clustered on state \times month of birth. Sample is limited to the subset of the original estimation sample born within four months after and four months before the minimum kindergarten entry age in their Wave 4 state of residence.

public kindergarten. I produce estimates of the impacts of attendance by substituting a smooth function of the same form as in Model 3 for the unrestricted eligibility effects in Model 2, then estimating that model using TSLS with *elig_{is}* as an instrument.

Estimates for the full sample of states considered in the main analysis are given in Table 8. Underlying models also include as controls the demographic and background variables in Table 1, Panel B, as well as age and dummies for month-by-year of assessment during the Wave 4 interview. There is a strong first-stage relationship between age-eligibility for kindergarten and public kindergarten attendance in 2006–2007—on average, a child just barely eligible was 66 percentage points more likely to be enrolled

in the full sample (Column 1).³⁷ The marginal child in the full sample also scored a significant 0.38 standard deviations higher on the kindergarten-age reading and math tests (Column 2). Point estimates are much more similar across income in this case, and the implied effect of public kindergarten attendance in the sample overall—57 percent of a standard deviation (Column 3)—is essentially the same as what I found for universal pre-K in the full sample (Table 2).³⁸

VII. Cost–Benefit Analysis

As laid out in Section II, per-pupil costs of the universal and targeted pre-K programs under study are similar. However, the attendance impact for the average child eligible for a universal program—approximately 0.6 standard deviations (Table 2)—greatly exceeds the attendance impact for the average (low-income) child eligible for a targeted program—approximately -0.08 standard deviations (Table 3). While the findings thus suggest a more favorable benefit-to-cost ratio for universal programs, it is useful to formalize this calculation.

Table 9, Panel A presents estimates of the benefit-to-cost ratio for each program type.³⁹ Only per-pupil state outlays for pre-K are available, so I make two conservative assumptions regarding total per-pupil spending that are well above the state contribution. The marginal social benefit of universal pre-K attendance, measured as the present discounted value (PDV) of the expected earnings gains from the increase in test scores, is 39 percent higher than the per-pupil cost of K–12 schooling (Panel A, Row 1). It is 85 percent higher than the per-pupil cost of K–12 schooling net of the social savings from substitution from other public and private centers. These ratios are not statistically different from one but are close to significantly different from those for targeted pre-K programs (Column 3), as seen in Column 4. Using Head Start to approximate program outlays (Panel A, Row 2) considerably increases these ratios for universal pre-K. The 3.52 benefit-to-cost ratio in the most favorable (yet still conservative) scenario for universal pre-K is both greater than one and greater than the corresponding ratio for targeted programs. It is also larger than the (highly significant) benchmark estimate for kindergarten programs, of 2.96 (Column 1).

Through it is customary to present benefit-to-cost ratios for early educational programs, the marginal value of public funds (MVPF)—or the ratio of a program beneficiary’s willingness to pay for the program out of their own income to government costs net of fiscal externalities—provides a means of using causal estimates from program evaluation for social welfare analysis (Hendren 2016). I thus also present estimates of the MVPF of each program, making a small adaptation to the MVPF formula used by Kline and Walters (2016) in their reevaluation of the Head Start Impact Study to accommodate the contemporaneous private benefits (to families) of universal programs. Specifically, the MVPF in the present context pins the marginal program beneficiary’s

37. Notably, this estimate is considerably lower than the kindergarten enrollment rate of age-eligible children. This is to be expected given the commonly found extent of noncompliance with entry-age regulations, but lack of information on exact birthday in the ECLS-B may attenuate this estimate further. I also find a low first-stage coefficient on the instrument for pre-K attendance relative to the levels of pre-K attendance reported by NIEER.

38. However, substitution from alternative care is also more common than in the universal pre-K case: 73 percent switch from other formal center-based care (Online Appendix Table 10).

39. See table notes and Online Appendix B for a description of assumptions over key parameter values.

Table 9
Cost-Benefit Analysis under Alternative Assumptions

	Universal K (1)	Uni. Pre-K (2)	Tar. Pre-K (3)	<i>p</i> : Uni. = Tar. (4)
Panel A: Benefit-to-Cost				
1. Program outlays = per-pupil K-12 spending Cost = program outlays	1.98 (0.45) [0.03]	1.39 (0.77) [0.61]	-0.18 (1.03) [0.25]	0.16
Cost = net program outlays	2.96 (0.68) [0.00]	1.85 (1.02) [0.41]	-0.24 (1.32) [0.35]	0.15
2. Program outlays = p.p. Head Start spending Cost = program outlays	n.a.	2.17 (1.20) [0.33]	-0.27 (1.50) [0.40]	0.14
Cost = net program outlays	n.a.	3.52 (1.94) [0.20]	-0.40 (2.20) [0.52]	0.12

(continued)

Table 9 (continued)

	Universal K (1)	Uni. Pre-K (2)	Tar. Pre-K (3)	p : Uni. = Tar. (4)
Panel B: Marginal Value of Public Funds (MVPF)				
1. Program outlays = per-pupil K–12 spending	3.94 (1.63) [0.07]	1.96 (1.39) [0.49]	−0.24 (0.99) [0.21]	0.14
2. Program outlays = p.p. Head Start spending	n.a.	4.27 (4.19) [0.43]	−0.39 (1.58) [0.38]	0.25

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in calendar year 2001, “Children’s Birth Certificates” (collection at nine months), “Parent–Guardian Interviews,” (collection at 48 months), “Direct Child Assessments,” “Early Care and Education Providers,” and “School Questionnaires” (collection at kindergarten entry).

Notes: Benefit-to-cost ratios and the MVPF calculated under two assumptions: per-pupil program outlays are equal to per-pupil K–12 spending (an upper bound for pre-K programs) and per-pupil program outlays are to per-pupil Head Start spending. For the benefit-to-cost ratio, benefits are calculated by multiplying the preferred IV estimates of the per-pupil attendance impact on test scores (with additional controls, low-income subsample in the case of targeted pre-K) by the predicted effect of a one standard deviation increase in early life test scores on lifetime earnings; net costs subtract fiscal savings from substitution from Head Start and from other private center-based care. For the MVPF, the beneficiary’s marginal willingness to pay sums predicted earnings impacts net-of-taxes and the transfer to parents from substitution from other private center-based care; net government costs subtract the fiscal externalities from substitution from Head Start and predicted tax revenues from per-pupil program outlays. Throughout, I assume a one standard deviation increase in test scores yields a 10 percent increase in earnings, that targeted pre-K attendees have lifetime earnings approximately 80 percent of those for the average child, and that the PDV of lifetime earnings at age 12 is on average \$522,000 (in 2010 dollars, from Chetty et al., 2011). I also assume a discount rate of 3 percent to calculate PDVs of lifetime earnings at age five (Column 1) and age four (Columns 2–3), a tax rate of 20 percent, and a per-pupil annual price of private center-based care equal to approximately \$5,000, based on Laughlin (2013) calculations from the Survey of Income and Program Participation. Per-pupil K–12 and Head Start spending is from Barnett et al. (2006) and corresponds to the 2005–2006 for the relevant states. Standard errors (in parentheses) are calculated using the delta method; p -values in brackets test of whether the ratio is equal to one.

willingness to pay to both their net-of-tax earnings gains from program participation and the reduction in their family's immediate out-of-pocket childcare expenses.

In particular, for program j , the MVPF is given by

$$MVPF^j = \frac{(1 - \tau)p \hat{\beta}_{TSLs}^j + \phi_j^V S_j^V}{\phi_j^j - \phi_j^B S_j^B - \tau p \hat{\beta}_{TSLs}^j},$$

where $\hat{\beta}_{TSLs}^j$ is the TSLs estimate of the effect of attending program j on test scores (in standard deviation units), p is the predicted change in discounted lifetime earnings with a one standard deviation test score increase, and τ represents the marginal tax rate. The first numerator term is thus the PDV of the net-of-tax private earnings gains from participation for the marginal program j attendee. The second is then the private transfer from program substitution—the product of the marginal private program cost, ϕ_j^V , and the likelihood of switching from a private center to program j , S_j^V . The denominator subtracts the marginal fiscal savings from substitution across public programs, $\phi_j^B S_j^B$, and the marginal discounted value of future tax revenues, $\tau p \hat{\beta}_{TSLs}^j$, from marginal government outlays for program j , ϕ_j^j . It thus captures the predicted cost to government of the marginal child's attendance, net of the fiscal externalities from public program substitution and the additional tax revenue.⁴⁰

The MVPF estimates in Table 9, Panel B are noisier than those for the benefit-to-cost ratios but show a similar pattern to the estimates in Panel A. Estimates of the MVPF are larger for universal pre-K and when program outlays are set equal to per-pupil Head Start spending. Indeed, in that case, the MVPF of universal pre-K is 4.27. This estimate is much higher than that seen for a number of redistributive policies (Hendren 2016) or even for Head Start (Kline and Walters 2016). However, I have more confidence in the conclusion that the MVPF of universal pre-K truly differs from that of targeted pre-K under the assumption of higher (K–12) costs.

But comparing the MVPFs of universal and targeted programs would only be truly meaningful for a welfare analysis if they served the same populations. Put differently, if these programs both only served low-income children, we could then draw the further conclusion that universal pre-K is the more desirable policy. While most of the true beneficiaries of universal pre-K appear to be low-income children, I cannot rule out that higher income children gain and so cannot draw this conclusion. Estimates of the MVPF for universal pre-K should therefore be helpful reference points for future work where monetary returns of universal pre-K attendance are measured in other data or perhaps more directly. By the same reasoning, however, my findings do not provide a strong case for targeted *state-funded* pre-K programs: The MVPF for targeted, state-funded pre-K programs is significantly lower than that for the federally funded Head Start program (Kline and Walters 2016), which serves essentially the same income group.

VIII. Conclusion

This work has presented new estimates of the impacts of preschool education in the United States and, by harnessing the benefits of observing children from across the country in an underused longitudinal survey, moved the literature

40. The benefit-to-cost ratios presented in Table 9 Panel A can be represented by this notation. The first, gross cost estimate is given by $p \hat{\beta}_{TSLs}^j / \phi_j^j$. The second, net cost estimate is given by $p \hat{\beta}_{TSLs}^j / (\phi_j^j - \phi_j^B S_j^B - \phi_j^V S_j^V)$.

forward in meaningful ways. First, I have presented comparable estimates of the immediate cognitive test score impacts of both universal and targeted pre-K, based not only on the same data but also the same research design. I have found evidence that the benefits from attending state-funded pre-K in a state with a universal program exceed those from attending state-funded pre-K in a state with a targeted program—not just overall but especially for the low-income children that both types of programs serve. I also attempted to rule out alternative explanations—beyond pre-K access—for the differences in pre-K attendance impacts across these two groups of states. Throughout, my empirical approach addressed limitations of prior pre-K evaluations exploiting age-eligibility rules for identification. Most notably, survey data from the ECLS-B supported estimation of valid TOT effects and incorporation of new tests of internal validity.

The constellation of evidence is consistent with universal pre-K delivering greater benefits to the population it serves, relative to the costs, than targeted pre-K. In other words, there may be an efficiency-related justification for choosing a universal program over a targeted one, at least in the context of state-funded pre-K. The fact that universal pre-K delivers short-term benefits similar to universal kindergarten suggests that it looks considerably more like public education than targeted pre-K. Public education being perhaps the one investment in children that has the most political support in United States, political economy considerations may be important for understanding universal pre-K's relative success.

This study has limitations. Though short-term cognitive test score gains from educational intervention predict impacts on adult outcomes like earnings (Chetty et al. 2011)—an idea that I use in the cost–benefit analysis—the findings in this paper only pertain to short-term cognitive effects for one birth cohort. Estimating the impacts of universal and targeted pre-K on longer-term outcomes, once the programs are mature enough to do so, would resolve uncertainty about whether the test score impacts documented here truly manifest in better outcomes over the longer term. Even looking at short-term outcome for more than one cohort would be helpful, if only because the present estimates are noisier than would be ideal. To my knowledge, however, the ECLS-B is the only data set currently available in the United States to carry out this analysis.

In addition, while the findings here provide a concrete example of a public program where universal eligibility may raise cost efficacy, the implications might not extend to situations where public goods are not directly provided by the government, or even beyond early education. Given the idiosyncratic features of policy debates over access, it is important to study the impacts of eligibility rules on the output and productivity of other public programs directly.

References

- Aizer, Anna, and Flavio Cunha. 2012. “The Production of Human Capital: Endowments, Investments and Fertility.” NBER Working Paper 18429. Cambridge, MA: NBER.
- Almond, Douglas, Janet Currie, and Valentina Duque. 2018. “Childhood Circumstances and Adult Outcomes: Act II.” *Journal of Economic Literature* 56(4):1360–446.
- Anders, John, Andrew Barr, and Alex Smith. 2021. “The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s.” *American Economic Journal: Economic Policy*. Forthcoming. <https://www.aeaweb.org/articles?id=10.1257/pol.20200660>.

- Anderson, Patricia, Kristin Butcher, Elizabeth U. Cascio, and Diane Schanzenbach. 2011. "Is Being in School Better? The Impact of School on Children's BMI When Starting Age Is Endogenous." *Journal of Health Economics* 30(5):977–86.
- Araujo, M. Caridad, Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *Quarterly Journal of Economics* 131(3): 1415–53.
- Araujo, M. Caridad, Marta Dormal, and Norbert Schady. 2019. "Child Care Quality and Child Development." *Journal of Human Resources* 54(3):656–82.
- Barnett, W. Steven, Allison H. Friedman-Krauss, G.G. Weisenfeld, Michelle Horowitz, Richard Kasmin, James H. Squires. 2017. *The State of Preschool 2016*. New Brunswick, NJ: The National Institute for Early Education Research.
- Barnett, W. Steven, Jason T. Hustedt, Laura E. Hawkinson, and Kenneth B. Robin. 2006. *The State of Preschool 2006*. New Brunswick, NJ: The National Institute for Early Education Research.
- Barnett, W. Steven, Kwanghee Jung, Allison Friedman-Krauss, Ellen C. Frede, Milagros Nores, Jason T. Hustedt, Carolee Howes, and Marijata Daniel-Echols. 2018. "State Prekindergarten Effects on Early Learning at Kindergarten Entry: An Analysis of Eight State Programs." *AERA Open* 4(2):1–16.
- Barr, Andrew, and Chloe R. Gibbs. 2022. "Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood." *Journal of Political Economy*. Forthcoming. <https://doi.org/10.1086/720764>
- Bitler, Marianne P., Hilary W. Hoynes, and Thurston Domina. 2014. "Experimental Evidence on Distributional Effects of Head Start." NBER Working Paper 20434. Cambridge, MA: NBER.
- Carneiro, Pedro, and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy* 6(4):135–73.
- Cascio, Elizabeth U. 2015. "The Promises and Pitfalls of Universal Early Education." *IZA World of Labor* 116.
- Cascio, Elizabeth U., and Ethan G. Lewis. 2006. "Schooling and the Armed Forces Qualifying Test: Evidence from School Entry Laws." *Journal of Human Resources* 41(2):294–318.
- Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." *Brookings Papers on Economic Activity*, Fall, 127–78.
- Cascio, Elizabeth U., and Douglas Staiger. 2012. "Knowledge, Tests, and Fadeout in Educational Interventions." NBER Working Paper 18038. Cambridge, MA: NBER.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4):1593–660.
- Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85(3):341–64.
- de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9):2964–96.
- Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3):111–34.
- Dickert-Conlin, Stacy, and Todd Elder. 2010. "Suburban Legend: School Cutoff Dates and the Timing of Births." *Economics of Education Review* 29(5):826–41.
- Elango, Sneha, Jorge Luis Garcia, James J. Heckman, and Andres Hojman. 2016. "Early Childhood Education." In *Economics of Means-Testing Transfer Programs in the United States*, Volume 2, ed. Robert Moffitt, 235–97. Chicago, IL: University of Chicago Press.
- Elder, Todd, and Darren Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources* 44(3):641–83.

- Engel, Mimi, Amy Claessens, and Maida A. Finch. 2013. "Teaching Students What They Already Know? The (Mis)Alignment between Mathematics Instructional Content and Student Knowledge in Kindergarten." *Educational Evaluation and Policy Analysis* 35(2):157–78.
- Feller, Avi, Todd Grindal, Luke Miratrix, and Lindsay Page. 2016. "Compared to What? Variation in the Impacts of Early Childhood Education by Alternative Care Type." *Annals of Applied Statistics* 10(3):1245–85.
- Fitzpatrick, Maria D. 2008. "Starting School at Four: The Effect of Universal Prekindergarten on Children's Academic Achievement." *B.E. Journal of Economic Analysis & Policy* 8(1):46.
- . 2010. "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Pre-kindergarten." *Journal of Labor Economics* 28(1):51–85.
- Fitzpatrick, Maria D., David Grissmer, and Sarah Hastedt. 2011. "What a Difference a Day Makes: Estimating Daily Learning Gains during Kindergarten and First Grade Using a Natural Experiment." *Economics of Education Review* 30:269–79.
- Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2015. "Integrated Public Use Microdata Series, Current Population Survey: Version 4.0 [Data Set]." Minneapolis, MN: University of Minnesota. <http://doi.org/10.18128/D030.V4.0>.
- Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4):999–1012.
- Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225(2):254–77.
- Gormley, William T., and Ted Gayer. 2005. "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program." *Journal of Human Resources* 40(3):533–58.
- 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):114–28.
- Hendren, Nathaniel. 2016. "The Policy Elasticity." *Tax Policy and the Economy* 30(1):51–89.
- Johnson, Rucker C., and C. Kirabo Jackson. 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy* 11(4):310–49.
- Kline, Patrick, and Christopher Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131(4):1795–848.
- Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2):497–532.
- Kuhfeld, Megan, and James Soland. 2021. "The Learning Curve: Revisiting the Assumption of Linear Growth across the School Year." *Journal of Research on Educational Effectiveness* 14(1):143–71.
- 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):162–87.
- Lipsey, Mark W., Dale C. Farran, and Kerry G. Hofer. 2016. "Effects of a State Prekindergarten Program on Children's Achievement and Behavior through Third Grade." Working Paper. Nashville, TN: Vanderbilt University, Peabody Research Institute.
- Lipsey, Mark W., Kerry G. Hofer, Nianbo Dong, Dale C. Farran, and Carol Bilbrey. 2013. "Evaluation of the Tennessee Voluntary Prekindergarten Program: Kindergarten and First Grade Follow-Up Results from the Randomized Control Design." Research Report. Nashville, TN: Vanderbilt University, Peabody Research Institute.
- 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.
- Puma, Michael, Stephen Bell, Ronna Cook, and Camilla Heid. 2010. *Head Start Impact Study Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Reardon, Sean. 2011. "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Changes*, ed. Greg J. Duncan and Richard J. Murnane, 91–116. New York: Russell Sage Foundation.
- Sabol, Terri J., Sandra L. Soliday Hong, Robert C. Pianta, and Margaret R. Burchinal. 2013. "Can Rating Pre-K Programs Predict Children's Learning?" *Science* 341:845–6.
- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- Thompson, Owen. 2018. "Head Start's Long-Run Impact: Evidence from the Program's Introduction." *Journal of Human Resources* 53(4):1100–39.
- Walters, Christopher. 2015. "Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start." *American Economic Journal: Applied Economics* 7(4):76–102.
- 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):2112–30.
- Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. 2008. "An Effectiveness-Based Evaluation of Five State Prekindergarten Programs." *Journal of Policy Analysis and Management* 27(1):122–54.