Economics of Education Review 31 (2012) 1143–1161

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Earnings benefits of Tulsa's pre-K program for different income groups

Timothy J. Bartik[a,*], William Gormley[b], Shirley Adelstein[b]

[a] W.E. Upjohn Institute for Employment Research, 300 S. Westnedge Ave., Kalamazoo, MI 49007-4686, United States
[b] Georgetown University, 37th & O Sts., NW, Washington, DC 20057, United States

ARTICLE INFO

Article history:
Received 24 March 2012
Received in revised form 23 July 2012
Accepted 31 July 2012

JEL classification:
I21
I24
J24

Keywords:
Economic impact
Efficiency
Rate of return

ABSTRACT

This paper estimates future adult earnings effects associated with a universal pre-K program in Tulsa, Oklahoma. These projections help to compensate for the lack of long-term data on universal pre-K programs, while using metrics that relate test scores to social benefits. Combining test-score data from the fall of 2006 and recent findings by Chetty et al. (2011) on the relationship between kindergarten test scores and adult earnings, we generate projections of adult earnings effects and a partial cost–benefit analysis of the Tulsa pre-K program. For both half-day and full-day programs, benefits are similar across program participants of different income, with benefit-to-cost ratios of 3- or 4-to-1. Because we only consider adult earnings benefits, actual benefit–cost ratios are likely higher, especially for disadvantaged children.

1. Introduction

As state pre-K programs, both targeted and universal, have blossomed throughout the United States, public officials have sought hard evidence on these programs' short-term and long-term effects. We now have substantial evidence on the short-term effects of targeted (Reynolds, Temple, Ou, Arteaga, & White, 2011; Schweinhart et al., 2005) and universal (Gormley, Phillips, & Gayer, 2008; Henry et al., 2003) programs and the long-term effects of targeted programs (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Reynolds, Temple, Ou, et al., 2011; Reynolds, Temple, White, Ou, & Robertson, 2011), but limited evidence on the long-term effects of universal programs (Karoly & Bigelow, 2005). Because the first universal pre-K program in the United States was not established until 1997, we will need to wait many years to estimate the consequences of universal pre-K for adults. Even when long-term data are available on universal pre-K, analysts will face the challenge that universal pre-K programs lack experimental data.

It is possible, however, to make some informed projections, using recent data from Oklahoma's universal pre-K program. Using information from recent work by Chetty et al. (2011) on the link between early test scores and adult earnings, we estimate future earnings effects of pre-K for children who were and were not eligible for a free school lunch. These projections, though not without their limitations, enable us to estimate some of the long-term benefits of a high-quality pre-K experience to children from different socioeconomic strata.

Using data collected on children who were beginning pre-K and kindergarten in Tulsa Public Schools in the fall of 2006, we use a regression-discontinuity design to estimate treatment effects on average test-score percentiles. We combine these results with Chetty et al.'s findings to generate projected earnings effects and a partial cost–benefit analysis of the Tulsa pre-K program. To make our Tulsa estimates consistent with Chetty's results, we use a different test score metric than in previous research on Tulsa pre-K (Gormley, 2010; Gormley & Gayer, 2005; Gormley, Gayer, Phillips, & Dawson, 2005; Gormley et al., 2008). To allow

* Corresponding author. Tel.: +1 269 343 5541; fax: +1 269 343 3308.
E-mail address: bartik@upjohn.org (T.J. Bartik).

0272-7757/$ – see front matter
http://dx.doi.org/10.1016/j.econedurev.2012.07.016

for benefit–cost analysis by income group and half-day versus full-day pre-K, we go beyond previous Tulsa results in segmenting estimates by both income group and half-day versus full-day program. Our analysis suggests that the Tulsa pre-K program has substantial earnings benefits for each of the income (school lunch) and program-type (full-day versus half-day) subgroups examined. In each case, the benefit–cost ratio well exceeds 1, even though we only consider adult earnings benefits in this analysis. We also find that the percentage effects on expected future adult earnings are largest for lower-income children.

In the next section, we discuss the limitations of test scores for assessing program benefits and making comparisons across income categories. We then describe recent findings by Chetty et al., which suggest the shape of the relationship between test-score percentiles and adult earnings. We describe the Tulsa pre-K program. We also describe our regression-discontinuity estimation technique. Next, we present program effects using average test score percentile as our outcome. We then use these estimated program effects to generate projected adult earnings benefits for different income and program-type subgroups. These projected earnings benefits serve as the basis for a partial cost–benefit analysis of the Tulsa pre-K program. Finally, we discuss the limitations of our analysis and offer some concluding remarks.

2. Measuring the distribution of preschool benefits: the need for a test score metric that can be related to benefits

2.1. Existing studies of preschool for different income groups

Should preschool be targeted at the disadvantaged, or universal? The answer depends on how benefits of preschool vary with a child's family income. Many benefits, such as higher adult earnings for former preschool participants, are long-term.

Long-term studies of preschool have focused on programs serving disadvantaged children, such as the Perry Preschool Program (e.g., Schweinhart et al., 2005), the Abecedarian Project (Campbell & Ramey, 2010), or the Chicago Child–Parent Center Program (Reynolds, Temple, Ou, et al., 2011).

Only one high-quality study of preschool for advantaged families has even medium-term follow-up. This study used random assignment to evaluate a preschool affiliated with Brigham Young University (Larsen & Robinson, 1989). With a modest sample size (125 treatment, 71 controls), they found some statistically significant effects of preschool on student achievement in 2nd and 3rd grades. However, this sample was unusual, with 97 percent two-parent families and 84 percent non-working mothers. In addition, this study does not include disadvantaged families, which prevents income group comparisons.

Some good short-term studies of preschool include children from both disadvantaged and more advantaged families (Gormley et al., 2005, 2008; Wong, Cook, Barnett, & Jung, 2008). The most rigorous short-term studies use a regression discontinuity methodology. This methodology, explained further below, allows us to estimate the effects of preschool on test scores at kindergarten entrance.

2.2. Metric challenges

Though test scores are often used to evaluate programs, test scores have limitations as a metric. In the present case, it is unclear how kindergarten test-score effects are related to program benefits.

The metric issue for test scores goes even further. Without further assumptions, kindergarten test-score effects do not allow qualitative statements comparing achievement gains across income groups. Without further information, test-score scales are arbitrary. Unless the individuals being compared have the same starting point, it is impossible to say who has gained the most.

Consider these hypothetical test-score effects of preschool. Suppose that children eligible for a free lunch go from a test score of 20 at preschool entrance to a test score of 40 at kindergarten entrance. Suppose that children who must pay full price for lunch go from a test score of 40 at preschool entrance to a test score of 60 at kindergarten entrance.

Which income group has gained the most from preschool to kindergarten? Without additional information, it is impossible to say. Each group has gained the same 20 points on this particular test-score metric. But we have no idea whether a gain from 20 to 40 points, versus 40 to 60 points, has the same benefits.

As Lang (2010) has pointed out, the arbitrariness of test-score metrics is a general problem for educational policy. For example, the arbitrariness of test-score metrics causes problems in comparing "value added" across teachers.

Comparing test-score gains requires relating test-score gains to something we value. We would like to relate test-score gains to at least an important component of total social benefits.

2.3. A possible solution

Fortuitously, recent research by Chetty et al. (2011) suggests the shape of the relationship between early test scores and adult earnings. Adult earnings gains are a major component of the benefits of many educational programs, including preschool. In the most recent benefit–cost analyses done of Perry Preschool (Heckman, Moon, et al., 2010), adult earnings gains are 51 percent of total social benefits and over 100 percent of the benefits for former program participants. (Former program participants lose welfare benefits.) In the most recent benefit–cost analysis of the Abecedarian program (Barnett & Masse, 2007), adult earnings gains of former child participants are 27 percent of total social benefits, and over 100 percent of the benefits to former child participants. In the most recent benefit–cost analyses of the Chicago Child–Parent Center Program (Reynolds, Temple, White, et al., 2011), adult earnings gains are 31 percent of total social benefits and 72 percent of the benefits to former program participants.[1]

Chetty et al.'s research uses data from the Tennessee Class-Size Study, also known as the Student/Teacher Achievement Ratio (STAR) experiment. This study was intended to examine the effects of class size in grades K-3. Students and teachers were randomly assigned to classrooms of different sizes.

Chetty et al.'s research links STAR participant data to adult earnings data from the IRS. Chetty et al. find a simple relationship between one test score metric at kindergarten exit and adult earnings. The Chetty et al. metric for test scores at kindergarten exit is the average percentile rank of the student on the various tests combined. This involves scoring the test as the student's percentile rank, using the control group sample or a national sample, and then taking the average of these percentile ranks. As mentioned by Chetty, this test-score metric has a precedent in Krueger (1999).

Chetty et al. relate this percentile metric to average adult earnings from age 25 to 27. Chetty et al. find that this percentile metric is linearly related to adult earnings (Chetty et al., 2011, Figs. I and IV). An increase in test scores of 10 percentiles gives approximately the same dollar increase in adult earnings, regardless of the initial percentile rank.

Chetty et al. examine the impact of class quality on test scores and adult earnings. Class quality is measured by average test scores of all students other than the individual student being considered. For kindergarten entrants, Chetty et al. find that an increase in kindergarten class quality that raises an individual student's test scores at the end of kindergarten by 1 percentile increases average annual adult earnings at ages 25–27 by $73.01 (Chetty et al., 2011, Appendix Table 13, col. 1).[2] This measure controls for the individual student's demographic characteristics. Because students were randomly assigned to classes in the STAR experiment, this estimate can be interpreted as causal. This causal estimate is somewhat lower than the raw correlation between kindergarten test scores and adult earnings. We use Chetty et al.'s estimate of test score effects on adult earnings to estimate the effects of Tulsa's pre-K program on adult earnings.

2.4. Potential drawbacks

Extrapolating Chetty et al.'s impact estimates to our data raises legitimate issues. First, it is not necessarily the case that early test-score impacts generated by policy *X* will yield similar effects on adult earnings to similar test-score impacts generated by policy *Y*. Perhaps the two policies will have different accompanying changes in students that will lead to different effects on adult earnings.

For example, the long-term effects of early childhood programs may be in part due to these programs' effects on "soft skills" (Heckman, Malofeeva, Pinto, & Savelyev, 2010; Heckman, Stixrud, & Urzua, 2006). These soft skills include social skills such as ability to get along with peers and teachers, as well as character skills such as self-confidence. Soft skills would be correlated with reading and math scores, but imperfectly so. Soft skill effects might differ between the "class quality" differences in Chetty et al. and the Tulsa pre-K program. We know that pre-K programs sometimes are associated with greater behavior problems (Magnuson, Ruhm, & Waldfogel, 2007).

However, as we will describe later, the evidence indicates that Tulsa's pre-K program improves social skills. Therefore, the Tulsa program may be likely to have relatively strong adult outcome effects compared to its test score effects.

We might also question whether the changes caused by these two programs (Tennessee STAR and Tulsa pre-K) arise through similar mechanisms. Similar mechanisms make it more plausible to extrapolate from one program to the other. The Chetty et al. estimates rely on the mechanism of exogenous changes in classroom quality. The Tulsa pre-K estimates might appear to rely on pre-K versus no pre-K.

However, as we will describe later, our study estimates the effects of Tulsa pre-K relative to whatever the pre-K entrants did during the preceding year. In most cases, this comparison group participated in some non-paternal care. Survey evidence from pre-K entrant families reveals that 56 percent of these children were in non-paternal care during the preceding year, 42 percent in center-based pre-K or child care. What distinguished Tulsa pre-K is that it offered higher classroom quality, as documented in a previous report (Phillips, Gormley, & Lowenstein, 2009). Therefore, the mechanisms of policy intervention are more similar between Chetty et al. and our Tulsa pre-K evaluation than might be immediately apparent.

A second issue is that our preschool impacts on test scores are estimated for the start of kindergarten, whereas Chetty et al.'s results link adult earnings to test scores at the end of kindergarten. However, Chetty et al. find that the estimated effect of test-score percentile on adult earnings is similar across grade levels from the end of kindergarten to the end of fourth grade (Chetty et al., 2011, Appendix Tables 4 and 5). This makes it plausible that the end of kindergarten/adult earnings relationship will be similar to the start of kindergarten/adult earnings relationship.

A third issue with extrapolating Chetty et al.'s results is differences in local labor markets. The link between early test scores and adult earnings in Tennessee might differ from the link in Tulsa. However, migration of even a minority of workers and employers across local labor markets should limit the differentials across local labor markets in the relationship between adult earnings and adult skills (Marston, 1985). If early childhood skills are linked to adult skills, then migration also puts some limits on how much the relationship between early childhood skills and adult earnings can differ across local labor markets.

[1] The Reynolds, Temple, White, et al. (2011) study updates a benefit–cost analysis previously presented in Temple and Reynolds (2007), which also provides comparative benefit–cost analyses across these three well-known programs.

[2] Chetty et al. (2011) measure effects in 2009 dollars as $78.71. However, in this paper, we measure all dollar effects in fiscal year 2005–06 values, e.g., from July 1, 2005, to June 30, 2006. This is done to later be comparable with the cost data in Tulsa, which is for the 2005–06 fiscal year.

In sum, there are good reasons to be skeptical that Chetty et al.'s results can provide predictions of pre-K's effects on adult earnings that are precise. Because our extrapolation from Chetty et al. to Tulsa has considerable uncertainty, our simulations consider a number of alternative scenarios for forecasting earnings effects. In addition, the next section considers whether Chetty et al.'s results do yield reasonable extrapolations for three pre-K programs for which we have actual measures of both early test-score effects and later adult outcomes.

2.5. Evidence from Ypsilanti, Chapel Hill, and Chicago

To see whether Chetty et al.'s results can forecast adult earnings gains from pre-K, we consider evidence from the Perry Preschool Program, the Abecedarian Program, and the Chicago Child–Parent Center Program. These three programs are the only high-quality preschool programs with good evidence on adult outcomes in the mid-20s or later.

For each program, we consider how the program affected test scores. These test-score effects are translated into percentile effects. Based on Chetty et al.'s results for dollar earnings effects and mean earnings at ages 25–27, we calculate a percentage effect on adult earnings.[3] Percentage effects on earnings are used because the preschool programs' estimated effects on adult earnings are at a variety of ages. It seems reasonable that over the life cycle, earnings effects will vary with baseline earnings.[4] In addition, the earnings metrics used in the studies differ, which suggests that percentage comparisons are more appropriate.[5] These test-score-predicted percentage effects on adult earnings are compared with direct estimates of these programs' effects on adult earnings, based on adult outcomes.

As Table 1 shows, the test-score-predicted adult earnings effects based on Chetty et al.'s results are reasonably close to the adult-outcome-predicted adult earnings effects. These findings suggest that Chetty's results can be extrapolated to some preschool programs.

Finally, there are precedents for using test scores to predict future earnings effects of early interventions. In Krueger's (2003) analysis of Tennessee STAR, he relied on Currie and Thomas's (1999) estimates of how reading and mathematics test scores at age 7 are related to adult earnings at age 33, using British data. The Currie and Thomas estimates of the test score/adult earnings relationship are similar to the Chetty et al. estimates. An analysis by Duncan, Ludwig, and Magnuson (2010) cited Krueger in using Currie and Thomas's estimates to provide what they described as "rough estimates" of the adult earnings benefits from Head Start and state government-funded pre-K programs.

Table 1
Comparison of percentage effects of preschool on adult earnings.

Program	% Earnings effects predicted from end of preschool test-score effects	% Earnings effects predicted from adult outcomes
Perry Preschool	16.0	19.4
Abecedarian Project	10.1	13.8
Child–Parent Center	7.8	7.3

Note: Perry Preschool test-score effects come from Schweinhart et al. (2005, Table 3, p. 61). We calculated average effect size at the end of the second preschool year for these tests: the Stanford-Binet IQ Test, the Leiter International Performance Test, the Peabody Picture Vocabulary Test, and the Psycholinguistic Abilities Test. Average effect size was then translated into change in percentiles. Change in percentiles was multiplied by the Chetty et al. (2011) estimates that implied that a 1-percentile change in test scores increases annual earnings by 0.495%. Perry Preschool percentage effects on adult earnings come from Heckman, Moon, et al. (2010, Table 3, p. 119). Percentage earnings gains were based on undiscounted gross earnings effects. Abecedarian Project test score effects are mean Weschler full-scale test score effects at age 60 months (Ramey and Campbell, 1991, Table 8.4). Abecedarian adult earnings effects are measured using educational attainment effects at age 31 reported in Campbell et al. (2012). We use these educational attainment effects to predict earnings effects using data from the Current Population Survey Outgoing Rotation Group on how employment rates, weekly hours, and wage rates for blacks differed by educational attainment at age 31. Chicago Child–Parent Center Program test scores come from Reynolds (1995, Table 3, p. 15). We calculated average effect size for the following kindergarten tests: cognitive readiness at kindergarten entry from the Iowa Tests of Basic Skills (ITBS), end of kindergarten reading readiness (ITBS), end of kindergarten math achievement (ITBS), and end of kindergarten teacher ratings of student's school adjustment. CPC percentage effects on adult earnings were calculated from Reynolds, Temple, Ou, et al. (2011, Table 2).

3. Estimating the percentile test-score effects of Tulsa's pre-K program

3.1. Overview

This section of the paper describe the Tulsa pre-K program and our data, outlines and defends our methodology for estimating test-score effects, and presents those test score effects.

3.2. Tulsa pre-K program

The Tulsa Public Schools pre-K program is a school-based, state-funded pre-K program for four-year-old children. Since 1998, Oklahoma's school districts have had the option of providing pre-K to all four-year-olds. Most school districts have participated, including Tulsa, the state's largest school district. Although enrollment is voluntary, most parents have enrolled their four-year-olds. Oklahoma now is second in the nation in preschool access, with over 70 percent of four-year-olds enrolled (Barnett, Carolan, Fitzgerald & Squires, 2012). During this time period, Tulsa enrolled around 60 percent of four-year-olds.

The Tulsa pre-K program is high quality. Every teacher has a B.A. degree, is early-childhood-certified, and earns

[3] Chetty et al.'s estimated $73.01 dollar increase in annual earnings is a 0.495 percent increase in annual earnings.

[4] Heckman, Moon, et al.'s (2010) analysis finds that over the life cycle, the percentage effects of Perry Preschool grow from the early 20s to the prime earning years. This is discussed further later in our paper. Thus, assuming that percentage effects will be constant is likely to be a conservative assumption about earnings effects.

[5] Chetty et al. (2011) use earnings data from IRS Form 1040 and Form W-2. Heckman, Moon, et al. (2010) use a comprehensive measure of compensation that includes benefits. Reynolds, Temple, Ou, et al. (2011) use a measure of median earnings within particular income groups. Our estimates of effects from the Abecedarian program are derived from program effects on educational attainment, as the earnings effects estimated are imprecise (Campbell et al., 2012).

the same salary as other public school teachers. A 10-to-1 child/staff ratio is maintained. A comparison of Tulsa Public Schools' pre-K classrooms with school-based pre-K classrooms in 11 other states reveals that Tulsa pre-K teachers spend more time on task than their counterparts elsewhere. Based on the CLASS measure, instructional quality is also higher in Tulsa than elsewhere (Phillips et al., 2009). The Tulsa program goes beyond a narrow focus on reading and math skills, stressing "concept development" and "feedback" more than other school-based programs. Previous analyses of the Tulsa pre-K program have found some statistically significant but modest positive effects on social-emotional development (Gormley, Phillips, Newmark, Welti, & Adelstein, 2011).

In short, the Tulsa pre-K program is a better-than-average pre-K program that reaches large numbers of students. In contrast, the justly celebrated Perry Preschool and Abecedarian Project programs only served small numbers of students. Tulsa pre-K is also relatively low cost. Whereas the Tulsa pre-K program's average cost per child was \$4403 (half-day)/\$8806 (full-day) (more on these estimates later), Perry Preschool cost \$17,526 per child, and the Abecedarian Project cost \$39,672 per child (Barnett & Masse, 2007, Table 1; Schweinhart et al., 2005, Table 7.8, p. 148; all figures in 2005–06 dollars).[6] Tulsa pre-K is similar in costs to the Chicago CPC program, which cost \$5372 per year for a half-day program.[7] The Chicago CPC is also larger scale than Perry or Abecedarian, though not as large as the Oklahoma program.

3.3. Tulsa data

Our data come from student testing in August 2006. Just prior to the commencement of classes, teachers administered three subtests of the Woodcock–Johnson Achievement Test to incoming kindergarten students and incoming pre-K students. The three subtests were (1) Letter-Word Identification (a measure of prereading skills), (2) Spelling (prewriting skills), and (3) Applied Problems (premath skills). These tests were successfully administered to 73 percent of incoming kindergarten students and 78 percent of incoming pre-K students.

We also obtained data from administrative records and from a parent survey, also conducted in August 2006. Administrative records specified each child's gender, date of birth, race/ethnicity, and school lunch status, a surrogate for income. The parent survey, received from 86 percent of tested students, yielded valuable information on the mother's education, the presence of the biological father at home, Internet access, and other variables.

In the analyses that follow, we handle missing data using multiple imputation (Little & Rubin, 2002; Rubin, 1987).[8] This method creates multiple complete data sets with plausible values for missing data based on observed values. The complete data sets are analyzed separately and combined to produce the final results, which incorporate the uncertainty associated with imputation. Parameter estimates are averages of estimates across the imputed data sets, and standard errors are calculated according to Rubin's (1987) method, which accounts for both within- and between-imputation variance. Multiple imputation has been shown to outperform other missing data techniques (e.g., Croy & Novins, 2005; Rubin, 1996; Sinharay, Stern, & Russell, 2001).

In contrast to previous Tulsa studies (Gormley & Gayer, 2005; Gormley et al., 2005, 2008), we have created a different dependent variable, to mirror Chetty et al. (2011). Previous Tulsa studies have utilized raw test scores and have reported results for the three subtests separately. In this study, we utilize percentile ranks and combine the three subtests into one measure. We scaled scores for each subtest into percentile ranks based on the universe of tested kindergarten students, regardless of treatment status. We then assigned percentile ranks for each subtest to both pre-K entrants and pre-K alumni and took the average across the three subtests.

3.4. Estimating technique

As with previous studies of Tulsa pre-K (Gormley & Gayer, 2005; Gormley et al., 2005, 2008), we used a regression-discontinuity design. This design has also been used by studies of other pre-K programs (Wong et al., 2008). We compared incoming kindergarten students who had participated in pre-K the previous year (the treatment group) with incoming pre-K students (the comparison group). This research design addresses the concern that certain families are more likely to select into the pre-K program, and these families may have unobservable characteristics that affect test scores.

Whether a child is in the treatment or comparison group depends on date of birth. The state of Oklahoma enforces a strict birthday cutoff for program eligibility. For the 2005–06 academic year, children were only qualified to attend the TPS pre-K program if they were born on or before September 1, 2001. Because of this strict birthday cutoff (the discontinuity), we can see whether test scores abruptly change between pre-K entrants who just missed the birthday cutoff versus pre-K alumni who just made it.

Obviously, the students in the pre-K entrant and pre-K alumni group differ in average age, which is positively related to test outcomes. However, our regression discontinuity approach controls for the students' age and other

[6] Perry Preschool and Abecedarian cost more in part because they covered multiple years: two years for most Perry participants, five years for most Abecedarian participants. Benefit and cost figures for these programs in most studies, including this one, pool all participants.

[7] Data from Reynolds, Temple, White, et al. (2011) but adjusted to 2005–06 prices. The CPC program provided services for two years. However, only 55 percent of program participants participated for two years. The one-year program had a higher benefit–cost ratio than the two-year program (Reynolds, Temple, White, et al., 2011, Table 5).

[8] We implemented multiple imputation with five imputes using the Stata ICE program (Royston, Carlin, & White, 2009), which generates plausible values using imputation by chained equations (Van Buuren, Boshuizen, & Knook, 1999). For our analytic sample, missing data was less than 1 percent for race, 20 percent for whether the child lived with his/her biological father, and 27 percent for mother's education. Increasing the number of imputes had a negligible impact on the estimates.

Table 2
Comparison of average test percentile and covariates for TPS pre-K entrants and alumni.

Variable	Pre-K entrants			Pre-K alumni			Diff.	*p*
	M	SE	*N*	*M*	SE	*N*		
Average test percentile	27.031	0.859	1418	41.897	1.327	1256	−14.866	0.000
Female	0.499	0.026	1418	0.482	0.029	1256	0.017	0.660
Race								
Black	0.325	0.025	1393	0.345	0.027	1252	−0.021	0.566
White	0.349	0.025	1393	0.340	0.027	1252	0.009	0.810
Hispanic	0.210	0.021	1393	0.220	0.023	1252	−0.010	0.752
Native American	0.102	0.016	1393	0.082	0.017	1252	0.020	0.388
Asian	0.014	0.006	1393	0.012	0.007	1252	0.002	0.862
Lunch status								
Free	0.628	0.025	1403	0.677	0.028	1254	−0.049	0.187
Reduced-price	0.146	0.018	1403	0.102	0.019	1254	0.045	0.085
Full-price	0.226	0.023	1403	0.221	0.025	1254	0.005	0.884
Mother's education								
No high school	0.189	0.023	1070	0.223	0.027	887	−0.034	0.333
High school	0.262	0.027	1070	0.263	0.030	887	−0.001	0.986
Some college	0.397	0.029	1070	0.370	0.034	887	0.027	0.545
College degree	0.152	0.020	1070	0.144	0.024	887	0.008	0.795
Lives with father	0.622	0.028	1152	0.589	0.031	994	0.033	0.433
Internet access	0.469	0.028	1164	0.491	0.032	1002	−0.021	0.619

Note: This table compares regression-based estimates of different variables at the birthday cutoff for TPS participation in the 2005–06 academic year (September 1, 2001). We compare pre-K entrants who just missed being in pre-K the previous year and pre-K alumni who just made being in pre-K the previous year.

characteristics. The big advantage of our approach is that both sets of students participated in the program. Both sets of parents chose to enroll their students in the program. This should reduce selection bias due to unobservable characteristics associated with program participation, such as parent or student motivation.

For the regression-discontinuity approach to provide unbiased estimates of pre-K effects, the key assumption is that unobservable characteristics affecting test scores only change smoothly with age, and do not jump abruptly at the age cut-off. Is this assumption true? We did a number of tests of whether this assumption is reasonable.

First, if unobserved student characteristics "jump" at the birthday cutoff, we might expect similar jumps at the cutoff in observable characteristics. The results in Table 2 show the differences in test scores and observable characteristics around the birthday cutoff. These estimates use a 12-month window around the cutoff, and allow for age to have an effect on each dependent variable that is linear, but with a different slope before and after the cutoff, and a discontinuous jump at the cutoff. No other covariates are in these regressions. The estimates in the table are predicted differences at the cutoff. With the exception of the proportion receiving reduced-price lunch ($p < 0.10$), all of the differences between pre-K entrants and alumni are nonsignificant. This lack of a jump at the cutoff in observable characteristics suggests that unobservable characteristics also do not jump at the cutoff.

Second, we might want to visually confirm that the data show only one distinct jump in test scores, at the birthday cutoff. If jumps in test scores were evident at other ages, this might cast doubt on the birthday cutoff jump being due to the year of experience in Tulsa pre-K.

Fig. 1, panel (a), illustrates the unique jump in test scores at the birthday cutoff date. Age is measured as number of days born before or after September 1, 2001, the birthday cutoff. To the left of the cutoff, average test scores are based on pre-K entrants who qualified for TPS pre-K in 2006–2007 based on their birthdays. These children are our comparison group, as they had selected into the program but were just beginning it when tested. To the right of the cutoff, average test scores are based on pre-K alumni who participated in TPS pre-K in 2005–2006. These children comprise our treatment group, as they had recently completed the program and were just beginning kindergarten in TPS when tested. The plotted data points reflect mean values for average test percentile by two-percentile bins of age. The figure shows relatively smooth changes in test scores with age either before or after the cutoff, as estimated by the regression line in the figure. At the cutoff, there is an abrupt jump, equal to the difference in where the two regression lines hit the cutoff.

Panels (b) and (c) show similar comparisons of test scores for alumni and entrants into full-day versus half-day pre-K. These data also show a unique jump in test scores at the cutoff.

Third, we might be concerned that parental decisions about enrollment in public pre-K are correlated both with the child's age in relation to the birthday cutoff, and with unobservable factors affecting test scores. For example, Duncan et al. (2010, footnote 10 on pp. 46–47) argue that among children who just missed the birthday cutoff, the most highly motivated parents might choose to enroll their kids in private pre-K programs and then private kindergarten. This selective enrollment would bias down test scores of pre-K entrants close to the cutoff, and bias upward estimates of pre-K's effects using regression discontinuity.

In response to this concern, we would expect such selective enrollment to cause a jump in observable student characteristics at the cutoff. But we see no such jump in observable characteristics at the cutoff, which casts doubt on the argument for major problems due to selective enrollment.

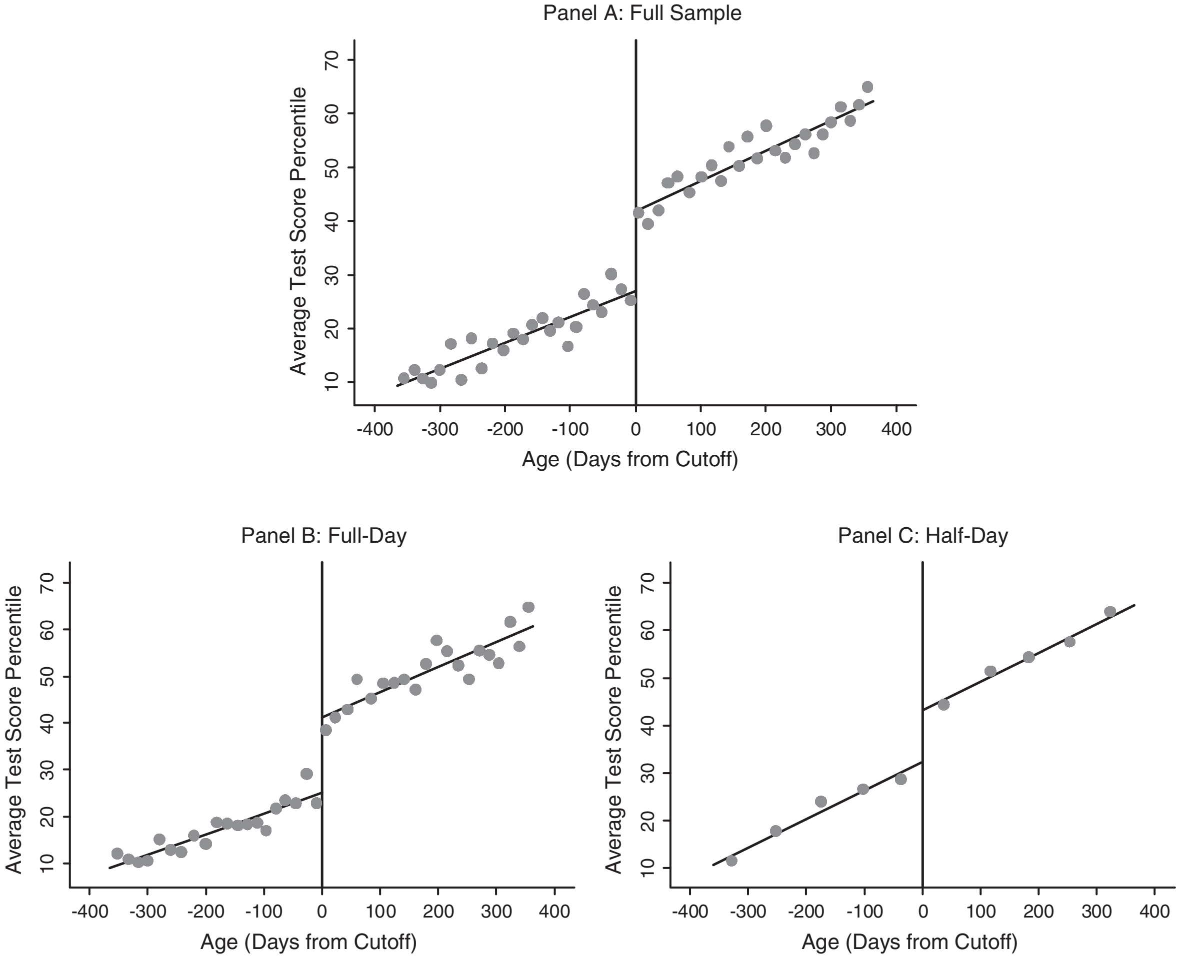

Fig. 1. Predicted average test score percentile before and after birthday cutoff. *Note*: These three panels show actual and predicted test scores for children in Tulsa pre-K. Panel A includes all children in Tulsa pre-K: panels B and C show full-day pre-K and half-day pre-K. Observations to the left of the cutoff show children who qualified for Tulsa pre-K in 2006–2007 based on their birthday. These children comprise our comparison group, because while they selected into the TPS pre-K program, they were just beginning the program at the time of testing. Observations to the right of the cutoff are children who qualified for and participated in TPS pre-K in 2005–2006. These children comprise our treatment group, as they had recently completed the program at the time of testing and were just beginning TPS kindergarten. The regression lines in each panel show predicted test scores for the comparison group and treatment group for separate regressions by group of test scores on age. The plotted data points reflect mean values for average test percentile for the "optimal bin size" for each sample. This bin size in each panel is "optimal" in that it passes two statistical tests suggested by Lee and Lemieux (2010): adding in twice as many bins yields an insignificant test on the added bins, and adding in interactions between the bin dummies and a linear term in age is not statistically significant. The optimal bin size is 2 percentile bins (e.g., about two weeks) for the full sample, 2.5 percentile bins for the full-day sample, and 10 percentile bins for the half-day sample.

We also note that private kindergarten enrollment in the Tulsa Public School district is only 12.7 percent of total kindergarten enrollment.[9] There is not much scope for effects due to selective private school enrollment associated with the age cutoff.

In addition, delayed kindergarten enrollment in the Tulsa Public Schools district is only 11 percent.[10] We exclude children from our estimation sample if they are the "wrong age" for either pre-K or kindergarten. Again, for this problem to occur there must be selective delay of children's enrollment that shows a jump at the cutoff, is correlated with test scores, and cannot be captured by observable characteristics.

We also looked to see if we could see wide variation in student enrollment in public pre-K and kindergarten by age. Appendix A shows that there are no obvious signs of gross variations in student enrollment by age in our sample.

Fourth, we might be concerned about selection bias because our pre-K alumni sample only includes students enrolled in Tulsa Public Schools. But we would expect major problems due to this selection of "stayers" to be reflected in a jump in observable characteristics at the cutoff, which is not evident in Table 2.

Fifth, we might be concerned about the sensitivity of results to how wide we make the window of observation

[9] Authors' calculations from pooled data from American Community Survey, 2005–2009.

[10] Authors' calculation using this study's data on kindergartners.

Table 3
Effect of TPS pre-K participation on average test score percentile.

	Full sample	Free lunch	Reduced-price lunch	Full-price lunch
Born before cutoff (treated)	15.535***	16.458***	15.843***	12.147***
	(1.586)	(1.795)	(4.697)	(3.102)
Age (days)	0.048***	0.039***	0.051***	0.073***
	(0.004)	(0.005)	(0.009)	(0.007)
Age*born before cutoff	0.003	0.015	0.000	−0.024*
	(0.007)	(0.009)	(0.023)	(0.013)
Black	−5.274***	−4.863***	−6.049**	−6.521***
	(1.199)	(1.567)	(2.766)	(1.946)
Hispanic	−7.050***	−7.946***	−8.112**	−0.485
	(1.203)	(1.590)	(3.196)	(2.953)
Native American	−0.959	−1.278	−2.067	0.185
	(1.316)	(1.694)	(2.709)	(2.667)
Asian	−0.941	−3.063	−1.499	0.953
	(2.209)	(3.929)	(8.141)	(3.753)
Female	5.923***	5.466***	5.644***	6.808***
	(0.722)	(0.859)	(1.970)	(1.258)
High school	2.900***	2.477**	4.623	6.920
	(1.011)	(1.033)	(2.908)	(4.784)
Some college	6.408***	6.080***	7.531**	11.081**
	(1.262)	(1.560)	(3.395)	(4.291)
College or higher	14.036***	8.836***	7.868**	21.323***
	(1.728)	(2.729)	(3.748)	(4.307)
Lives with father	1.254*	0.439	1.358	4.041*
	(0.721)	(0.880)	(2.668)	(1.968)
Internet access	4.812***	3.889***	4.376*	8.459***
	(0.733)	(0.844)	(2.374)	(2.135)
Free lunch	−6.514***			
	(1.315)			
Reduced-price lunch	−5.607***			
	(1.484)			
Constant	23.928***	17.537***	18.882***	16.720***
	(1.709)	(1.914)	(4.284)	(4.152)
Observations	2645	1641	341	663
Effect size[a]	0.90	1.22	0.98	0.56
Percentage effect[b]	84	114	83	43
R-squared	0.56	0.54	0.54	0.55

Note: Outcome variable is the average test-score percentile across three Woodcock–Johnson achievement tests: Letter-Word ID, Spelling, and Applied Problems. Percentiles for each test are based on the distribution of test scores in the full, age-appropriate kindergarten sample. Robust standard errors adjusted for clustering by school are in parentheses. *R*-squared is averaged across imputes.

[a] Effect size is the treatment effect divided by the standard deviation of the outcome for the comparison group (pre-K entrants) in the relevant sample.

[b] Percentage effect is 100 times the treatment effect divided by the mean of the outcome for the comparison group (pre-K entrants) in the relevant sample.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

around the cutoff. We have chosen to use a 12-month bandwidth, to maximize our sample size and precision. In Appendix B, we report estimates using narrower bandwidths. We find that narrowing the age bandwidth does not yield statistically significantly different estimates or change our conclusions, but it does come at a sizable cost in precision.

In addition, we experimented with moving from a linear specification in age to quadratic, cubic, and quartic specifications. These more general specifications did not yield statistically significantly different results, but had an even greater cost in statistical precision. Appendix B also gives more detail on these functional form tests.

Sixth, we might be concerned with problems with the standard errors if misspecification leads the disturbance term to be correlated for students with the same birthdate. This problem is noted by Lee and Card (2008). Their suggested solution for the present case would be to correct the standard errors for clustering by birthdate. Appendix C explores this possibility. We find that clustering by birthdate makes no significant or substantive difference to our results.

3.5. *Preliminary results for comparison with previous studies*

Table 2 and Fig. 1 do not control for student characteristics other than age. To get estimates comparable to previous studies, we estimate a single equation model that includes several student characteristics as covariates. In this model, test scores are regressed on precise date of birth (number of days born before or after the cutoff), an indicator for whether the person was born before the cutoff, an interaction term between these two variables to allow for different slopes on either side of the cutoff, and other student characteristics. Table 3 presents the results of such a model, for the full sample and three lunch status subgroups: free-lunch, reduced-price lunch, and full-price lunch.

Table 4
Effects of TPS pre-K participation on average test score percentile, by lunch and full-day status, and for all income groups by full-day status.

	Full-day pre-K program				Half-day pre-K program			
	Free lunch	Reduced-price lunch	Full-price lunch	All income groups	Free lunch	Reduced-price lunch	Full-price lunch	All income groups
Treatment effect	18.132***	20.236***	16.549***	17.937***	11.951***	8.764	10.075**	10.727***
	(2.025)	(5.347)	(4.691)	(1.729)	(2.974)	(8.439)	(4.189)	(2.824)
Effect size[a]	1.37	1.32	0.72	1.07	0.84	0.48	0.50	0.59
Percentage effect[b]	130	112	54	103	72	41	38	50
Observations	1286	226	297	1809	354	114	366	834
R-squared	0.56	0.62	0.49	0.56	0.50	0.46	0.61	0.57

Note: Outcome variable is average test score percentile across three Woodcock–Johnson achievement tests: Letter-Word ID, Spelling, and Applied Problems. Percentiles for each test are based on the distribution of test scores in the full, age-appropriate kindergarten sample. Robust standard errors adjusted for clustering by school are in parentheses. R-squared is averaged across imputes. All regressions include the full set of control variables reported in Table 3 and linear specification for age; full results are available on request.

[a] Effect size is the treatment effect divided by the standard deviation of the outcome for the comparison group (pre-K entrants) in the relevant sample.

[b] Percentage effect is 100 times the treatment effect divided by the mean of the outcome for the comparison group (pre-K entrants) in the relevant sample.

** $p < 0.05$.

*** $p < 0.01$.

Similar to the results in Table 2, for the overall sample, program participation increases test scores by 15.5 percentile points on average, an effect size of 0.90. This effect size is slightly higher than in previous work on Tulsa pre-K, which examined the three tests separately and used raw test score rather than percentiles (e.g., the average effect size is 0.69 in Gormley, Phillips, Adelstein, & Shaw, 2010).

As one shifts along the income spectrum from full-price-lunch to reduced-price-lunch to free-lunch students, effect sizes shift steadily upward, from 0.56 to 0.98 to 1.22. These "effect size" variations across income groups are qualitatively similar to previous results for the Tulsa pre-K program (e.g., Gormley et al., 2005 found effect sizes of 0.49 for the full-price-lunch group, versus 0.64 for the free-lunch group).

Another metric for comparing different income groups is to examine percentage effects on test scores compared to mean test scores at pre-K entrance. As shown in Table 4, percentage effects are 43 percent for the full-price-lunch group, increasing to 83 percent for the reduced-price-lunch group, and then up to 114 percent for the free-lunch group. Much of this variation is due to the unsurprising fact that low-income children at pre-K entrance tend to have much lower test scores on average. These percentage variations are qualitatively similar to previous Tulsa results (Gormley, 2010).

However, neither "effect sizes" nor "percentage effects" are linearly tied to the benefits of pre-K programs. Based on Chetty et al., we should instead be focusing on the gain in test scores measured in percentile terms, which is linearly tied to the predicted gains in adult earnings.

3.6. *Percentile test score effects disaggregated by lunch status and half-day versus full-day program*

We focus more attention in this subsection on the percentile effects, which will allow later forecasts of adult earnings gains. To allow for a comparison of adult earnings benefits with program costs, we must disaggregate the Tulsa data by whether students were in a half-day or full-day program. Obviously, the full-day program costs more. This further disaggregation is particularly important because lower-income groups were much more likely to be enrolled in full-day pre-K. Of the free-lunch group enrolling in pre-K, 78 percent were in full-day (and hence 22 percent in half-day). For the full-price-lunch enrolling in pre-K, 45 percent were in full-day pre-K (and 55 percent in half-day). (The reduced-price lunch group in pre-K was in-between, with 66 percent in full-day, 34 percent in half-day.) This disaggregation of results simultaneously by both full-day versus half-day program, and by income group, is an important difference between this study and previous Tulsa studies.

These different enrollment patterns across income groups are related to differences in what programs are offered in different neighborhood schools. The state of Oklahoma provides a greater state subsidy for full-day pre-K than for half-day pre-K, and also a greater subsidy for pre-K provided to students eligible for a free lunch. In addition, local school districts can use federal Title I funds to help support pre-K for students eligible for a free or reduced-price lunch. Finally, low-income families may have greater need for full-day programs, both to help their children and to provide free child care. Therefore, Tulsa chooses to concentrate more of its full-day pre-K programs in neighborhood schools that serve a higher proportion of disadvantaged children.

In addition, the different enrollment patterns may be related to differences in parent preferences. Parents are not allowed to select a full-day versus a half-day program at a particular school, as each school only offers one program type. However, parents choose whether to participate in the program at all. In addition, there were two Tulsa public schools in 2005–06 that offered some full-day slots to nonneighborhood children, with free transportation. Low-income parents may find full-day pre-K programs more appealing, while middle-class parents may be more attracted to half-day programs.[11]

[11] Conversations with Tulsa school officials suggest that most parents prefer full-day pre-K programs. In response, as of the 2010–11 school year,

Enrollment differences across income groups must be controlled to compare the benefits and costs of pre-K across income groups. We do not want to conclude that pre-K is more effective for low-income groups simply because a higher percentage enrolls in full-day programs.

Table 4 presents the percentile test-score effect of Tulsa pre-K, disaggregated by full- versus half-day status, and then also by lunch status. We only present the estimated effects of pre-K, but the full regression includes all of the control variables included in Table 3.

For a given type of program (full-day versus half-day), percentile test effects show no significant differences, statistically or substantively. But if were to rely on effect-size measures or percentage measures, we would conclude that Tulsa pre-K has effects on test scores for free-lunch children, versus full-price-lunch children, that are in some sense larger.

However, based on Chetty et al.'s results, the effects on test-score percentiles may be more directly related to a main benefit of pre-K, the effects on adult earnings. We now turn to forecasting adult earnings effects in order to do a rough benefit–cost analysis of the Tulsa pre-K program for different income groups.

4. Predicting the future earnings benefits of the Tulsa pre-K program

4.1. Overview

This section uses the estimated effects of Tulsa pre-K on the average test-score percentile for different income groups, combined with Chetty et al.'s estimates, to predict future earnings effects, which are then compared with program costs.

4.2. Predicted future earnings benefits

Even without much number-crunching, it can easily be seen that our Tulsa results and Chetty et al.'s results imply large forecasted effects on future earnings. The Chetty et al. results imply an annual earnings effect at ages 25–27 of around $73 per a 1-percentile boost. Given that this would probably go up as this cohort achieves higher earnings, and given that this earnings boost would accrue over a 40-year career, even a 1-percentile boost to test scores would have a large present value. Given that the Tulsa pre-K program boosted the percentile test scores by 9–20 percent, the implied effects are quite large.

To predict adult earnings effects, we used microdata from the American Community Survey (ACS) for the years 2005 through 2007 for the Public Use Microsample Areas (PUMAs) that correspond to the Tulsa metro area. We calculated annual earnings by age level from ages 22 to 66. (These ages are as of the survey date, which ranges through the year; the earnings data is over the 12 preceding months.) We started with age 22 to minimize complications due to effects of pre-K on educational attainment. We ended with age 66 to minimize complications due to mortality. Earnings data were adjusted to fiscal year 2005–06 prices.

We then calculated the effect of a 1-percentile early test-score boost at each age. Our baseline assumption was that the dollar effect of a 1-percentile boost would vary at different ages with the average earnings level at each age. We scaled the average earnings effects of a 1-percentile boost so that the percentage earnings effects match Chetty et al.'s estimates for ages 25–27.

This assumption of a constant percentage effect at different ages is likely to be conservative. Heckman, Moon, et al.'s (2010) analysis of the Perry Preschool results suggests that the percentage effects of preschool on former participants' adult earnings tend to increase at later ages. In sensitivity analysis, to be described below, we also considered alternative forecasts for how adult earnings effects varied with age.

We then discounted earnings gains to age 4, to be comparable later with Tulsa pre-K program costs. Our baseline results used a real discount rate of 3 percent, which is a fairly typical social discount rate. However, because we did not include any appreciation in real wages in these calculations, the true discount rate is really 3 percent plus whatever annual real earnings increase ends up occurring between age 4 and the relevant age.

With the baseline assumptions, we end up with a discounted present value of an increase of $1502 in adult earnings for a 1-percentile increase in test scores. Multiplying this $1502 figure by the estimates in Table 4 for the percentile effects across the six groups (full-day versus half-day times the three income groups) yields the present value for each group of these earnings increases. These make up the first row of numbers for Table 5.

As expected, these estimated earnings benefits of Tulsa pre-K are large. These earnings benefits would justify program costs much greater than what Tulsa pre-K actually costs. (We consider costs further a little later.) Furthermore, based on the similar effects on average test percentile for the different income groups, the predicted adult earnings effects do not differ much across income groups.

These earnings effects are roughly similar to the one previous attempt we know of to use kindergarten test scores to estimate the earnings effects of state pre-K programs. Using the estimates of Krueger (2003) and Currie and Thomas (1999), the study by Duncan et al. (2010) estimated that state pre-K programs might increase the present value of future adult earnings by $13,034. The estimated effects of Tulsa pre-K tend to be somewhat higher, which is consistent with the Tulsa program's relatively high quality.

4.3. Predicting earnings effects for different income groups in percentage terms

Although from an efficiency perspective the earnings benefits for the different income groups are similar, society may put a greater weight on gains for lower-income groups. To compare earnings gains with expected income,

virtually every Tulsa school now offers only a full-day program. However, in the 2005–06 school year, from which our data come, 31.7 percent of the students are in half-day programs.

Table 5
Predicted effects of Tulsa pre-K on future adult earnings and ratios to costs, by lunch and full-day status.

	Full-day pre-K program			Half-day pre-K program		
	Free lunch	Reduced-price lunch	Full-price lunch	Free lunch	Reduced-price lunch	Full-price lunch
Present value of adult earnings increase ($)	27,242	30,404	24,864	17,955	13,167	15,137
Estimated present value of child's parents' earnings, as % of Tulsa metro average	34.1	67.3	135.3	34.1	67.3	135.3
Extrapolated present value of child's baseline future earnings, as % of Tulsa metro average	65.0	85.4	112.8	65.0	85.4	112.8
Baseline present value of child's future earnings, discounted back to age 4 ($)	261,308	343,038	453,430	261,308	343,038	453,430
Predicted percentage effect on child's present value of future earnings	10.4	8.9	5.5	6.9	3.8	3.3
Program costs ($)	8806	8806	8806	4403	4403	4403
Ratio of program earnings benefits to costs, baseline assumptions and baseline scenario	3.09	3.45	2.82	4.08	2.99	3.44
Ratio of program earnings benefits to costs, fixed dollar effect on annual earnings	2.22	2.50	2.01	2.89	1.97	2.71
Ratio of program earnings benefits to costs, percentage effects on earnings increase with age	8.20	9.22	7.43	10.65	7.25	9.98
Ratio of earnings benefits to costs at 5% discount rate	1.56	1.74	1.42	2.06	1.51	1.73
Internal rate of return (%)	6.4	6.8	6.1	7.3	6.3	6.7

Note: Assumptions are described in more detail in text. All except last two rows assume 3% real discount rate. The present value of adult earnings increases are derived from the treatment effects on kindergarten test scores reported in Table 4. Predicted percentage effects are calculated from previous table entries, dividing the present value of the predicted earnings increase by the present value of predicted future earnings. Ratios of program earnings benefits to costs are calculated from previous table entries on these items. The baseline scenario assumes fixed percentage effects on earnings with age. Alternative scenarios assume fixed dollar effects, or percentage effects that increase with age. The next to last row recalculate the benefit–cost ratios under the assumption of a 5% real discount rate, but returns to the baseline scenario of fixed percentage effects on earnings. The last row of the table calculates the internal rate of return for each day-length and income group: the maximum discount rate at which measured program benefits (in this case, only adult earnings) still exceed estimated program costs, under the baseline scenario.

we estimate baseline future earnings for children in each income group.

To predict future earnings, we isolated records from the American Community Survey of all children ages 4–18 who lived in the city of Tulsa and who attended public schools. (Restricting the sample to children age 4 would create too small a sample size.) We separated the children into three groups: those with family income less than or equal to 130 percent of the poverty line, who should be eligible for a free-lunch subsidy; those with family income greater than 130 percent but less than or equal to 185 percent of the poverty line, who should be eligible for a reduced-price lunch; and those with family income greater than 185 percent of the poverty line, who will have to pay full price for lunch.

For each group, we looked at parents' earnings. We calculated for each group the mean earnings of parents, broken down by gender and age. These mean earnings calculations used the child weights in the ACS. We ended up with the following sample sizes: for free-lunch children, 349 mothers and 152 fathers; for reduced-price-lunch children, 158 mothers and 114 fathers; and for full-price-lunch children, 494 mothers and 457 fathers.

Earnings for each of these six groups of adults, broken down by income group, gender, and age, were compared with the earnings of all adults in the Tulsa metro area broken down by gender and age. For each of the income-group/gender/age combinations, we calculated mean earnings as a percentage of mean earnings for all adults in the Tulsa metro area broken down by gender and age. We then calculated for each of the six groups the average percentage that parents' earnings are of Tulsa metro-area earnings, using the number of parents in each of the ages as weights. Finally, we averaged across mothers and fathers to get an average percentage ratio of each group's income to average overall metro-area earnings.

For the parents of children in Tulsa public school free-lunch families, average earnings were 34.1 percent of the average Tulsa metro-area earnings. For the parents of children in Tulsa public school reduced-price-lunch families, average earnings were 67.3 percent of average Tulsa metro-area earnings. For the parents of children in Tulsa public school full-price-lunch families, average earnings were 135.3 percent of average Tulsa metro-area earnings.

These are typical earnings for parents. We would expect some regression to the mean for the child's future earnings. The child's expected future earnings should be between his or her parents' earnings and overall Tulsa metro-area earnings.

The research literature suggests that a plausible value for the coefficient on ln(parent earnings), in a regression explaining the natural log of the child's earnings as an adult, might be 0.4 (Chadwick & Solon, 2002; Solon, 2002). As a result, the natural log of the child's average earnings will be a weighted average of the natural log of the parents' earnings and the natural log of overall average earnings, with a 0.4 weight on parent earnings.

Based on this calculation, we would expect the children of Tulsa free-lunch families to have average earnings as adults that are 65.0 percent of the overall Tulsa metro-area average ($65.0 = \exp[(0.4) * \ln(34.1) + (1 - 0.4) * \ln(100.0)]$). The children of Tulsa reduced-price-lunch families would be predicted to have average adult earnings of 85.4 percent of the overall Tulsa average. The children of Tulsa full-price-lunch families would be predicted to have average adult earnings of 112.8 percent of the overall Tulsa metro-area average.

The present value as of age 4 of Tulsa metro-area average earnings from ages 22 to 66, discounted at 3 percent, is $401,833. Using the above percentages, the expected present value of future earnings for the children from the three income groups can be calculated: free-lunch children, $261,308 (=65.0 percent × $401,833); reduced-price-lunch children, $343,038; and full-price-lunch children, $453,430. These baseline earnings figures can be compared with the dollar effects of Tulsa pre-K to get percentage effects.

These percentage effects are reported in the fifth row of numbers in Table 5. These percentage effects differ considerably across income groups. For example, for a half-day pre-K program, the percentage effects on earnings are over twice as great for the free-lunch group as for the full-price-lunch group (6.9 percent versus 3.3 percent). These differences are sufficient that a policymaker concerned with income inequity might reasonably conclude that the social benefits of pre-K are considerably greater for children from free-lunch families than for children from full-price-lunch families.

However, these percentage differentials are not as great as some might have guessed. This occurs because the child's future adult earnings tend to "regress to the mean." The free-lunch-eligible group's parents have earnings that are only about one-fourth of those of the parents of the full-price-lunch children (34.1 percent of the metro-area average versus 135.3 percent). But the expected future earnings of the free-lunch children are 58 percent of the expected future earnings of full-price-lunch children (65.0 percent of the metro-area average versus 112.8 percent of the metro-area average).

This illustrates the difficulty of targeting based on the child's future earnings. We can easily target pre-K and other interventions based on parental earnings. However, some children of the low-income parents will do considerably better as adults than their parents, even without any program intervention, and some children of the above-average income parents will do considerably worse as adults than their parents. The long-term nature of early childhood interventions creates difficulties in targeting precisely based on the future-needs status when the benefits of these interventions will be realized. Nevertheless, the preceding analysis still suggests that Tulsa pre-K's benefits may be disproportionately larger for participants from more disadvantaged families.

4.4. A partial benefit–cost analysis

These predicted future earnings effects can be used to do a partial benefit–cost analysis of Tulsa pre-K. The benefit cost analysis is partial because it does not include all benefits. Among the most important omitted benefits are reductions in crime and special education costs.

A benefit–cost analysis requires a reasonable estimate of Tulsa pre-K program costs. We estimated state aid to Tulsa pre-K by applying the state aid formula to the number of pre-K children enrolled and to those demographic characteristics of the student body that triggered additional increments of state aid, such as school lunch eligibility and English language learner status. We then added in federal Title I funds used for Tulsa pre-K. From conversations with Tulsa Public Schools officials, a rough estimate of the local funds used in Tulsa Public Schools in 2005–06 is 87 cents in local funds for every dollar of state aid. More details on cost estimates are in Appendix D.

This local share exceeds the estimated local share for all of Oklahoma for 2005–06 that is provided by the National Institute for Early Education Research (NIEER). NIEER estimates for 2005–06 that for each dollar of state aid for pre-K in Oklahoma, local school districts provide 57 cents (Barnett, Hustedt, Hawkinson, & Robin, 2006). We chose the larger local-share number for two reasons. First, we wanted to err on the side of overestimating program costs. Second, we felt it plausible that a large urban school district such as Tulsa might provide a larger local share than is typical in Oklahoma.

Conversations with Tulsa Public Schools officials also suggested that there were no differences in pre-K spending for children from different income groups. More low-income children were in full-day pre-K, as indicated previously, but for a given type of program, there are believed to be no systematic differences in spending across different income groups. Furthermore, Tulsa Public Schools officials felt strongly that the costs of full-day pre-K were simply twice the costs of half-day pre-K, so that there were no economies of scale from doubling the pre-K day.

We concluded that in 2005–06, a half day of Tulsa pre-K cost $4403, and a full day of Tulsa pre-K cost twice as much, at $8806. This cost estimate includes all program costs, whether from federal, state, or local dollars.

Combining these cost estimates with the estimated earnings benefits, we can come up with a partial benefit–cost ratio for the Tulsa pre-K program for different income groups and full-day versus half-day programs. These benefit–cost numbers are provided in the seventh row of numbers in Table 5.

For all income groups and all program dosages, the ratio of earnings benefits to costs is much greater than one. Either program length for each of the three income groups would pass a benefit–cost test and have net economic efficiency benefits. These conclusions would only be strengthened if other benefits were added in.

Within each program type, full-day versus half-day, benefit–cost ratios do not differ much between the free-lunch group and the full-price-lunch group. The benefit–cost ratios differ somewhat for the reduced-price-lunch group, but these estimates are less precise, because of the much smaller sample sizes for the reduced-price-lunch group (see Table 4). From a pure efficiency perspective, and if we include only predicted adult earnings benefits, there is not much reason to prefer additional pre-K services to low-income children over additional services to middle-class children. From an equity perspective, of course, there is a reason to prefer expanding services to low-income children. In addition, if the benefits of reduced crime or reduced special education costs, which are not measured here, were added, such benefits might differ across income groups.

It is tempting to use the benefit–cost figures in Table 5 to calculate the incremental benefits versus costs of a child moving from a half-day program to a full-day program or vice versa. This temptation should be resisted. Such calculations require the assumption that the incremental child switched from one program to the other would experience the same test-score effects and earnings benefits as the average child observed in that income group and program option. However, this may not be the case. Families participate in full-day versus half-day program types based partly on their own choices and partly on decisions by Tulsa Public Schools about where to place programs. The calculated incremental benefits versus costs of switching programs do not have the "selection correction" advantages of the regression-discontinuity method, which addresses selection bias by comparing treatment and comparison groups who were similarly selected into the program. A second important caveat is that full-day versus half-day programs may have additional benefits and costs. For example, offering full-day pre-K may make it easier for some families, particularly low-income families, to participate in the pre-K program.

4.5. Sensitivity to alternative assumptions

How sensitive are these results to alternative assumptions? First, consider the sensitivity to alternative percentile test score effects. The model is linear in percentile test score effects. The dollar effects and benefit–cost ratios change proportionately with percentile test score effects. Because the benefit–cost ratios are all more than two, even if we cut all the percentile test score effects in half, the benefit–cost ratios for all program types and groups would still exceed one.

Second, consider the sensitivity of these results to the Chetty et al. estimate of earnings effects at ages 25–27. The model is also linear in this parameter. Even if Chetty's results over-estimate how Tulsa pre-K affects earnings, the over-estimate would have to be extreme for benefit–cost ratios not to exceed one.

Third, consider the sensitivity of these results to our assumption that the percentage earnings effects at ages 25–27 will be the same with age over the life cycle. The model is not linear in this assumption. We consider two alternative scenarios.[12] One conservative scenario is that the annual dollar effects on earnings estimated by Chetty et al. apply at all ages from age 22 to age 66. This is a conservative scenario because dollar earnings effects at prime earnings years probably exceed dollar earnings effects at ages 25–27. This conservative scenario's effects on benefit–cost ratios are shown in Table 5, eighth row. Even under this conservative scenario, benefit cost ratios are 1.97 or greater for all income groups and program combinations.

A second alternative scenario is that percentage effects on earnings increase with age as estimated by Heckman, Moon, et al. (2010) for the Perry Preschool Program.[13] This scenario's implications for benefit–cost ratios are shown in Table 5, ninth row. Under this scenario, benefit cost ratios increase to seven or more for all groups.

Fourth, consider the sensitivity of these results to assumed discount rates. The model is not linear in discount rates. Therefore, the last rows of Table 5 reports results of some alternative assumptions. The next-to-last row returns to baseline assumptions, except for increasing the assumed real discount rate from 3 percent to 5 percent. Although this considerably reduces the benefit–cost ratios, these ratios are still greater than one.

We also calculated the rate of return to Tulsa pre-K. This is the maximum real discount rate at which the benefits still exceed costs. As the last row shows, these rates of return vary between 6.1 percent and 7.3 percent. These are quite healthy returns.

These rates of return are not far below the 7–10 percent rates of return for Perry Preschool estimated by Heckman, Moon, et al. (2010). This is remarkable, as the returns estimated by Heckman et al. also include the benefits of lower crime. Tulsa has relatively high returns because it is cheaper than Perry Preschool, yet achieves almost as great test score effects.

5. Conclusion

It is important to stress our findings' limitations. First, we cannot be sure that kindergarten test scores in Oklahoma will translate into adult earnings in precisely the same way that kindergarten test scores in Tennessee have translated into adult earnings. Different job markets could have different returns to different skills. Long-term effects of better kindergarten class quality in Tennessee could be more or less durable than the long-term effects of a pre-K program in Tulsa. If so, we could be underestimating or overestimating long-term effects.

[12] It is certainly good practice to consider alternative scenarios for how earnings will evolve in response to a shock in an unknown future, as is done in the pre-K context by Heckman, Moon, et al. (2010) or in a different labor market context by Oreopoulos, von Wachter, and Heisz (2012).

[13] Heckman, Moon, et al. (2010) estimated the following percentage earnings effects: ages 19–27, 7.9 percent; ages 28–40, 25.6 percent; ages 41–65, 21.1 percent. We assumed that Chetty et al.'s estimate of a 0.495 percent increase in earnings for a one percentile test score increase applied at ages 19–27, and used the ratios of Heckman et al.'s effects at different ages to project percentage effects at ages 28–40 and 41 and older.

Second, we have focused on earnings, while ignoring effects on remedial education and crime. Other studies of the long-term effects of pre-K have found crime effects to be important (Heckman, Moon, et al., 2010; Reynolds, Temple, Ou, et al., 2011; Rolnick & Grunewald, 2003; Schweinhart et al., 2005). This omission means that we are significantly underestimating long-term benefits. Furthermore, we are likely to be understating long-term benefits associated with pre-K for lower-income children relative to the long-term benefits for middle-income children.

Third, we have assumed that the long-term effects of pre-K can be predicted by short-term effects on literacy and math skills without including explicit recognition of the possible long-term influence of short-term effects on social-emotional development. This could be seen as assuming that long-term effects of pre-K are largely due to effects on literacy and math skills rather than effects on social-emotional development, which seems a strong assumption.

Our focus on short-term "hard skills" can be defended as assuming that short-term effects on hard skills can proxy for overall short-term effects. For example, better hard skills will lead a child to have more self-confidence, one of the soft skills. We do know, from Chetty et al.'s results and other results (e.g., Currie & Thomas, 1999) that short-term effects on hard skills can help to predict adult earnings, although the mechanism for transmitting these effects may not be solely through effects on hard skills. In Tulsa, we have witnessed both enormous cognitive gains and modest social-emotional gains in the short run. It is possible, however, that short-term improvements in soft skills have more profound long-term consequences than short-term gains in hard skills. A fuller understanding of the links between all these short-term effects and long-run earnings could lead to improved estimates of long-run earnings effects.

These reservations aside, our analysis offers some plausible estimates of future earnings effects for a high-quality pre-K program. It also illuminates benefit and cost differentials across income groups. What is most striking about these benefit–cost comparisons is not the modest variations by income groups but rather the similarities. For all children, irrespective of income, the earnings-related benefits alone of a high-quality pre-K program outweigh the costs by 3-to-1 or 4-to-1. This is an impressive accomplishment and one that should generate continued interest in the Tulsa pre-K model. Furthermore, although the results could be lower with more conservative assumptions about parameter estimates and discount rates, or more conservative scenarios for earnings paths, it would take sizable differences in assumed parameters or scenarios for the benefits of Tulsa pre-K not to exceed costs.

Our results also suggest that children from more disadvantaged families are likely to see the largest relative adult earnings benefits. Although benefit–cost ratios are similar across income groups, when adult earnings benefits are considered in percentage terms in relation to adult earnings prospects for each income group, sizable differences emerge. Free-lunch children have predicted percentage gains that are over twice as large as full-price-lunch children in half-day programs, a differential that is even larger for children in full-day programs. Furthermore, the similarity of benefit–cost ratios across groups reflects in part that this analysis has been limited to adult earnings benefits. Thus, the actual social benefits of the pre-K program in relation to program costs are likely to be even higher, particularly for disadvantaged children.

The results of this study speak to the potential for early childhood interventions to yield substantial benefits for children across socioeconomic strata. Although we recognize its limitations, we also believe that we have produced plausible estimates of the adult earnings benefits associated with the Tulsa pre-K program. Additionally, we have emphasized the importance of metrics for assessing program benefits that allow for meaningful comparisons across income groups by relating outcomes such as test-score gains to valued social benefits. In the case of the Tulsa pre-K program, the evidence suggests that with respect to at least one of these benefits – adult earnings – the results may be quite impressive.

Appendix A.

A.1. Enrollment in sample and in Tulsa pre-K by age

Fig. A1 reports the frequency in enrollment of children in our sample by age. Although the frequency is somewhat lower for pre-K alumni than pre-K entrants, this is due to a somewhat larger sample of pre-K entrants than alumni, due in part to student mobility out of the Tulsa school district. As mentioned in the paper's text, there are no obvious signs of gross variations in student enrollment by age in our sample.

In addition, although we do not have data on students in private kindergartens, we can look at students in public kindergarten, and see how the probability of student enrollment in the public pre-K program varied with birthdate. Table A1 reports these results. We do not see statistically significant differences in enrollment probabilities with birthdates. In particular, there is no statistically significant difference in the probability of enrollment in pre-K for students born in August, just before the cutoff, versus students born in September, just after the cutoff.

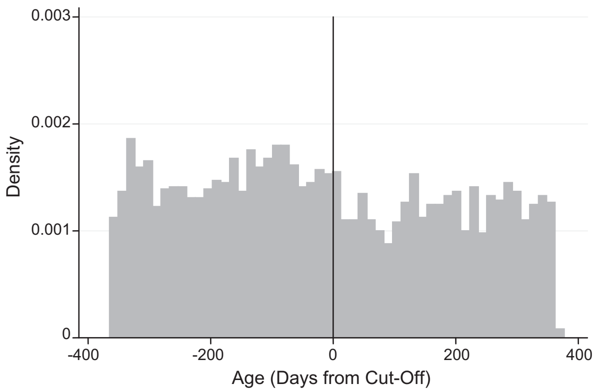

Fig. A1. Frequency of children in sample by age. *Note*: This figure shows the probability density of the children in the sample by age, with the probability density measured in two-week bins. Those children to the left are pre-K entrants, and those to the right are pre-K alumni.

Table A1
Regressions estimating probability of TPS pre-K program participation.

Month dummy for birth month	Coefficient (standard error)
September	0.049
	(0.043)
October	0.015
	(0.043)
November	0.021
	(0.041)
December	0.017
	(0.043)
January	0.008
	(0.042)
February	0.047
	(0.043)
March	0.034
	(0.042)
April	0.050
	(0.043)
May	−0.002
	(0.043)
June	−0.043
	(0.043)
July	0.060
	(0.043)

Note: This table reports a linear probability model for being in Tulsa pre-K the previous year among all Tulsa entering kindergartners who were not in Head Start. A logit model yielded similar results. Sample size was 3131. The model included as controls all the student socioeconomic characteristics and parent characteristics in Table 2. The mean probability in this sample of being in Tulsa pre-K the previous year was 50.0%. August was the omitted dummy. The month variables were not jointly significant, with a probability of an *F*-test of this size of 0.59.

Appendix B.

B.1. Sensitivity of estimated percentile test score effects to different bandwidths for estimation and different functional forms for effects of age

The baseline model in the paper includes all students in the pre-K entrant and pre-K alumni group who are within 12 months of the birthday cutoff. Age is assumed to have a linear relationship to test score performance, with this linear effect varying before and after the cutoff.

We considered alternatives to this specification. We considered narrowing the bandwidth. Table B1 reports estimated effects of pre-K on the average test score percentile using our baseline bandwidth of 12 months, and for three narrower bandwidths: 9 months; 6 months; 3 months.

Estimates become more imprecise as one narrows the bandwidth, as this throws out observations. The estimates also move around a bit. However, the resulting estimates do not differ statistically significantly from the estimates with the 12-month bandwidth.

The most robust finding is the strong effect of full-day pre-K for students eligible for a free lunch. Regardless of bandwidth, the estimated effects on this group are highly statistically significant. Effects of full-day pre-K for students eligible for a reduced-price lunch also are always statistically significant, although the magnitude moves around a bit more. In contrast, the estimates for the half-day pre-K program, which are smaller to begin with, tend to become insignificant if one moves to the narrowest bandwidth.

In an attempt to overcome imprecision problems with sample size, we used the estimates in Table B1 to calculate average effects segmented by income group, and by full-day versus half-day program. Table B2 shows the results. The income group comparisons combined the free-price and reduced-price lunch groups. To isolate the effects of income group, the income group comparisons weights each specification in Table B1 by the average percentage in the overall sample in full-day versus half-day programs, which is 68.4 percent in full-day programs, and the other 31.6 percent in half-day programs. To isolate the effects of full-day versus half-day, the full-day versus half-day averages hold the mix of income groups in each day length constant at the overall sample average, which is 62.1 percent free-lunch, 12.9 percent reduced-price-lunch, and 25.1 percent full-price-lunch.

Because of larger sample sizes, the average effects in Table B2 are somewhat more precise. The finding that percentile effects on low-income groups and middle-income groups are of similar magnitude is robust to sample bandwidth. In addition, the finding that students selected into the full-day program have larger effects than students selected into the half-day program is also robust to sample bandwidth.

We also ran all these window specifications in three alternative specifications that allowed the age effect to take on some non-linear form, with different parameters before and after the birthday cutoff. These nonlinear forms included quadratic, cubic, and quartic specifications. The imprecision of these estimates rapidly increased as the age effect was allowed to be more generally non-linear. Estimated effects in these specifications did not in general differ statistically significantly from linear specifications. In the 3-month specification, the Akaike Information Criterion preferred the linear functional form to all these non-linear specifications. This shorter bandwidth seems to incorporate any advantages that might be gained by non-linear specifications.

For illustrative purposes, Table B3 shows the quadratic specifications. Moving to the quadratic specification, especially when combined with shorter bandwidths, tends to drive up standard errors considerably. However, in general the coefficient estimates do not vary significantly from a linear specification with a 12-month bandwidth. One exception is the effect for reduced-price lunch students in a full-day program, which increases considerably in shorter bandwidths and a quadratic specification. However, the standard errors on some of these estimates are quite large.

To increase precision, Table B4 does a similar exercise to Table B2, but for the quadratic specification. Using the Table B3 estimates, we calculate average effects by income group, and average effects by full-day versus half-day program. The conclusion that different income groups have similar percentile test score effects, and that full-day effects are greater than half-day effects, seems robust to different quadratic specifications.

As a final attempt to overcome problems with sample precision, we look back again at results using the full

Table B1
Effects of TPS pre-K participation on average test score percentile, by lunch and full-day status, using different bandwidths.

Bandwidth	Full-day pre-K program			Half-day pre-K program		
	Free lunch	Reduced-price lunch	Full-price lunch	Free lunch	Reduced-price lunch	Full-price lunch
12 months						
Treatment effect	18.132***	20.236***	16.549***	11.951***	8.764	10.075**
	(2.025)	(5.347)	(4.691)	(2.974)	(8.439)	(4.189)
Effect size[a]	1.37	1.32	0.72	0.84	0.48	0.50
Percentage effect[b]	130	112	54	72	41	38
Observations	1286	226	297	354	114	366
R-squared	0.56	0.62	0.49	0.50	0.46	0.61
9 months						
Treatment effect	18.388***	18.920***	12.548*	11.848**	11.535	6.500
	(2.161)	(5.773)	(5.859)	(4.094)	(9.245)	(4.218)
Effect size	1.33	1.16	0.53	0.81	0.60	0.31
Percentage effect	123	95	37	63	47	21
Observations	980	168	218	273	89	262
R-squared	0.50	0.55	0.45	0.44	0.37	0.55
6 months						
Treatment effect	15.552***	22.440**	15.986*	8.451	7.804	4.907
	(2.692)	(8.463)	(8.296)	(5.401)	(12.661)	(5.522)
Effect size	1.07	1.28	0.66	0.56	0.39	0.22
Percentage effect	93	100	43	41	28	15
Observations	661	116	155	184	58	175
R-squared	0.44	0.44	0.38	0.37	0.48	0.45
3 months						
Treatment effect	12.540***	31.046***	19.267	5.731	3.597	11.687
	(2.923)	(9.324)	(11.399)	(6.047)	(17.592)	(10.084)
Effect size	0.74	1.76	0.78	0.35	0.17	0.47
Percentage effect	62	38	50	23	12	33
Observations	329	62	74	99	29	78
R-squared	0.34	0.49	0.27	0.27	0.43	0.48

Note: Outcome variable is average test score percentile across three Woodcock–Johnson achievement tests: Letter-Word ID, Spelling, and Applied Problems. Percentiles for each test are based on the distribution of test scores in the full, age-appropriate kindergarten sample. Robust standard errors adjusted for clustering by school are in parentheses. R-squared is averaged across imputes. All regressions include the full set of control variables reported in Table 4 and linear specification for age; full results are available on request.

[a] Effect size is the treatment effect divided by the standard deviation of the outcome for the comparison group (pre-K entrants) in the relevant sample.

[b] Percentage effect is 100 times the treatment effect divided by the mean of the outcome for the comparison group (pre-K entrants) in the relevant sample.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table B2
Average effects of pre-K for different income groups, and for full-day versus half-day, linear specifications with different bandwidths.

Bandwidth		Free and reduced price lunch group average	Full-price lunch group average	Full-day average	Half-day average
12 months	Coeff.	16.256***	14.506***	18.005***	11.070***
	St. error	1.589	3.472	1.854	2.385
9 months	Coeff.	16.370***	10.639**	16.991***	10.466***
	St. error	1.832	4.225	2.124	2.998
6 months	Coeff.	14.086***	12.490**	16.547***	7.479*
	St. error	2.405	5.939	2.882	3.975
3 months	Coeff.	12.451***	16.875**	16.608***	6.951
	St. error	2.712	8.426	3.592	5.060

Note: Average results reported here are weighted averages based on estimates reported in Table B1. The averages by income group assume that each income group participates in full-day versus half-day pre-K by the all-group average, which is 68.4 percent in full-day and 31.6 percent in half-day. The proportion of free lunch versus reduced price lunch students is assumed to be the same as the 12-month average, which is 82.8 percent versus 17.2 percent. The averages by full-day versus half-day assume the mix of students by income group in each day length of program is the same as the overall 12-month sample average, which is 62.1 percent free-lunch, 12.9 percent reduced-price-lunch, and 25.1 percent full-price lunch. Standard errors for averages are calculated based on standard errors in Table B1.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table B3
Quadratic specifications for effects of pre-K on average test score percentile, by lunch and full-day status.

Specification			Full day			Half day		
Bandwidth	Polynomial		Free	Reduced	Full price	Free	Reduced	Full price
12 months	Quadratic	Coeff.	16.561***	19.280**	12.523*	9.370*	10.050	9.774
		St. error	2.685	7.559	7.571	5.632	8.987	6.557
9 months	Quadratic	Coeff.	12.041***	25.004***	20.944**	6.661	8.325	9.682
		St. error	3.223	9.528	9.209	5.506	13.033	8.147
6 months	Quadratic	Coeff.	11.992***	37.234***	13.859	4.775	3.286	14.596
		St. error	3.543	10.914	10.817	6.593	15.158	10.191
3 months	Quadratic	Coeff.	12.356**	47.074***	13.187	6.609	−8.057	8.607
		St. error	5.009	13.503	20.728	9.575	40.934	12.736

Note: These estimates are for specifications identical to those in Table B1, except age is allowed to have quadratic effect on test scores, with parameters on age allowed to vary before and after the birthday cutoff.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table B4
Average effects of pre-K for different income groups, and for full-day versus half-day, quadratic specifications with different bandwidths.

Bandwidth		Free and reduced price lunch group average	Full-price lunch group average	Full-day average	Half-day average
12 month	Coeff.	14.648***	11.656**	15.898***	9.559**
	St. error	2.347	5.579	2.707	4.032
9 month	Coeff.	11.957***	17.391**	15.942***	7.633*
	St. error	2.676	6.807	3.292	4.320
6 month	Coeff.	12.601***	14.092*	15.707***	7.047
	St. error	3.053	8.072	3.764	5.203
3 month	Coeff.	13.829***	11.742	17.031***	5.224
	St. error	4.665	14.745	6.302	8.558

Note: Average results reported here are weighted averages based on estimates reported in Table B3. The averages by income group assume that each income group participates in full-day versus half-day pre-K by the all group average, which is 68.4 percent in full-day and 31.6 percent in half-day. The proportion of free lunch versus reduced price lunch students is assumed to be the same as the 12-month average, which is 82.8 percent versus 17.2 percent. The averages by full-day versus half-day assume the mix of students by income group in each day length of program is the same as the overall 12-month sample average, which is 62.1 percent free-lunch, 12.9 percent reduced-price-lunch, and 25.1 percent full-price lunch. Standard errors for averages are calculated based on standard errors in Table B3.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

sample. We consider how results vary for a linear specification when we move to very short bandwidths. Such local linear regressions can approximate any arbitrary functional form.

The results are shown in Table B5. Even if we move to narrow bandwidths on either side of the cutoff, estimates for the full sample do not vary much and are still highly statistically significant.

Table B5
Local linear regression results for different bandwidths for full sample, effects of pre-K on average test score percentile.

Bandwidth	Effects on test score percentile	Standard error	Sample size
12 months	15.513***	1.585	2643
9 months	15.523***	1.804	1990
6 months	14.062***	2.211	1349
3 months	13.692***	2.662	671
1 month	16.920***	5.079	240

*** $p < 0.01$.

Appendix C.

C.1. Sensitivity of standard errors to clustering by birthdate

Our estimates allow for clustering by school. However, the analysis of Lee and Card (2008) suggests we should also consider clustering by birthdate.

The rationale for clustering by birthdate is that there may be misspecification of the relationship between age and test scores. If there is such a misspecification, then the disturbance term includes the expected effect of this misspecification, which will then be correlated across individuals with the same birthdate.

Table C1 presents results using this different clustering. We just show results for the variable of interest, the discontinuous jump in test scores at the cutoff. As the table shows, clustering by birthdate makes little change to the estimated standard errors. In our particular case, changing the clustering to birthdate also changes the coefficient estimates, because the sample size slightly enlarges when adding the relatively few observations for which we do not have data

Table C1
Comparison of RD treatment effects using standard errors clustered by school versus date of birth.

Sample	School clustering	Date of birth clustering
Full sample	15.514***	15.535***
	(1.494)	(1.586)
Free lunch	16.352***	16.458***
	(1.887)	(1.795)
Reduced-price lunch	15.714***	15.843***
	(4.563)	(4.697)
Full-price lunch	12.380***	12.147***
	(3.233)	(3.102)
Full day	17.856***	17.937***
	(1.942)	(1.729)
Half day	10.832***	10.727***
	(2.549)	(2.824)
Full day, free lunch	18.020***	18.132***
	(2.226)	(2.025)
Full day, reduced-price lunch	20.236***	20.236***
	(5.510)	(5.347)
Full day, full-price lunch	16.549***	16.549***
	(5.007)	(4.691)
Half day, free lunch	11.793***	11.951***
	(3.636)	(2.974)
Half day, reduced-price lunch	9.004	8.764
	(8.162)	(8.439)
Half day, full-price lunch	10.440**	10.075**
	(4.103)	(4.189)

Note: Table shows estimated effects on test percentile, with standard errors in parentheses. Point estimates differ slightly due to the slightly larger sample size for the results that do not rely on school data. This is because a few observations are missing school data.
** $p < 0.05$.
*** $p < 0.01$.

on school assignment. However, this sample change also has few effects on coefficient estimates.

Appendix D.

D.1. Funding of Tulsa public schools pre-K program, 2005–06

We determined costs by combining relatively hard numbers on state and federal aid with softer estimates of local contributions.

STATE AID – We determined state aid by applying the state aid formula to students with different characteristics. The state aid formula took student eligibility for a school lunch and English language learner status into account. We determined the number of students eligible for a free or reduced price lunch from administrative data and used our parent survey to estimate the number of English language learners (our parent survey revealed that 14 percent of this cohort primarily spoke Spanish at home). The total state aid figure was probably slightly higher, because of 121 special education students. Without accurate information on how disabled each of these 121 students was, we cannot specify an additional increment in state funding.

TPS CONTRIBUTION – According to TPS (Joe Stoeppelwerth, personal correspondence, 6/21/11), local school revenue for the 2005–06 year was approximately 87 percent of state aid that year. We assume that local support for pre-K mirrored local support for the school system as a whole. An alternative estimate, by Barnett et al. (2006, p. 192), for Oklahoma as a whole, yields a different, lower level of local support.

FEDERAL AID – TPS supplied the federal aid figure for 2005–06 (Zelia Banks, personal correspondence, 5/6/11). Federal funding, from Title I of the ESEA, went to the ECDC program, the largest TPS pre-K provider at the time.

OTHER – For Oklahoma as a whole, "other" aid (e.g., private foundation support) is 3 percent of state aid (Barnett et al., 2006, p. 192). We used this same percentage to estimate other aid for the TPS pre-K program.

Total TPS pre-K spending for 2005–06 was $13,068,494 for 1788 students, 66 percent of whom were in full-day programs, with 34 percent in half-day programs. From conversations with TPS officials, we believe that TPS spent twice as much for each full-day student as for each half-day student. We conclude that full-day pre-K programs served 1180 children at $8806.26 per child and that half-day pre-K programs served 608 children at $4403.13 per child. We have no reason to believe that TPS spent more on free-reduced lunch students than students ineligible for a school lunch subsidy, even though the state aid formula took school lunch eligibility into account.

Financial support by funding source	
STATE AID	$6,686,896
TPS SPENDING	$5,817,600
FEDERAL AID	$363,391
OTHER	$200,607
TOTAL	$13,068,494
Costs by type of student	
Full-day (1180 students × $8806.26 per student):	$10,391,386.80
Half-day (608 students × $4403.13 per student):	$2,677,103.04
TOTAL EXPENSES	$13,068,489.84[a]

[a] The estimate of total expenses differs slightly from the estimate of total funding support because of rounding error.

References

Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. (2012). *The state of preschool 2011*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Hustedt, J. T., Hawkinson, L. E., & Robin, K. B. (2006). *The state of preschool 2006*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., & Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review*, *26*, 113–125.

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., et al. (2012). Adult outcomes as a function of an early childhood educational program: An abecedarian project follow-up. *Developmental Psychology* (advance online publication).

Campbell, F. A., & Ramey, C. T. (2010). Carolina abecedarian project. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life: A human capital integration* (pp. 76–98). New York: Cambridge University Press.

Chadwick, L., & Solon, G. (2002). Intergenerational income mobility among daughters. *American Economic Review*, *92*(1), 335–344.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics*, *126*(4), 1593–1660.

Croy, C. D., & Novins, D. K. (2005). Methods for addressing missing data in psychiatric and developmental research. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(12), 1230–1240.

Currie, J., & Thomas, D. (1999). *Early test scores, socioeconomic status, and future outcomes*. NBER Working Paper No. 6943. Cambridge, MA: National Bureau of Economic Research.

Duncan, G., Ludwig, J., & Magnuson, K. (2010). Child development. In P. B. Levine, & D. J. Zimmerman (Eds.), *Targeting investments in children fighting poverty when resources are limited* (pp. 27–58). Chicago: University of Chicago Press.

Gormley, W., Phillips, D., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development*, *82*, 2095–2109.

Gormley, W. T., Jr. (2010). Small miracles in Tulsa: The effects of universal pre-k on cognitive development. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life: A human capital integration* (pp. 188–198). New York: Cambridge University Press.

Gormley, W. T., Jr., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. *Journal of Human Resources*, *40*(3), 533–558.

Gormley, W. T., Jr., Gayer, T., Phillips, D. A., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41*(6), 872–884.

Gormley, W. T., Jr., Phillips, D. A., Adelstein, S., & Shaw, C. (2010). Head Start's comparative advantage: Myth or reality? *Policy Studies Journal*, *38*(3), 397–418.

Gormley, W. T., Jr., Phillips, D. A., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, *320*(5884), 1723–1724.

Heckman, J. J., Malofeeva, L., Pinto, R., & Savelyev, P. A. (2010). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the high/scope Perry preschool program. *Journal of Public Economics*, *94*(1–2), 114–128.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). *Report of the findings from the early childhood study: 2001–02*. Atlanta: Andrew Young School of Policy Studies, Georgia State University.

Karoly, L. A., & Bigelow, J. H. (2005). *The economics of investing in universal preschool education in California*. Santa Monica, CA: RAND Corporation.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*(2), 497–532.

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, *113*(485), F34–F63.

Lang, K. (2010). Measurement matters: Perspectives on education policy from an economist and school board member. *Journal of Economic Perspectives*, *24*(3), 167–182.

Larsen, J. M., & Robinson, C. C. (1989). Later effects of preschool on low-risk children. *Early Childhood Research Quarterly*, *4*, 133–144.

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*, 655–674.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281–355.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley and Sons.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51.

Marston, S. T. (1985). Two views of the geographic distribution of unemployment. *Quarterly Journal of Economics*, *100*(1), 57–79.

Oreopoulos, P., von Wachter, T., & Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, *4*(1), 1–29.

Phillips, D. A., Gormley, W. T., Jr., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-k programs. *Early Childhood Research Quarterly*, *24*(3), 213–228.

Ramey, C. T., & Campbell, F. A. (1991). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A. C. Huston (Ed.), *Children in poverty: Child development and public policy* (pp. 190–221). New York: Cambridge University Press.

Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly*, *10*(1), 1–31.

Reynolds, A. J., Temple, J. A., Ou, S. R., Arteaga, I. A., & White, B. A. B. (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science*, *333*(6040), 360–364.

Reynolds, A. J., Temple, J. A., White, B. A. B., Ou, S. R., & Robertson, D. L. (2011). Age 26 cost–benefit analysis of the child–parent center early education program. *Child Development*, *82*(1), 379–404.

Rolnick, A., & Grunewald, R. (2003). Early childhood development: Economic development with a high public return. *Federal Reserve Bank of Minneapolis's the Region*, *17*(4), 6–12.

Royston, P., Carlin, J. B., & White, I. R. (2009). Multiple imputation of missing values: New features for MIM. *Stata Journal*, *9*(2), 252–264.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The high/scope Perry preschool study through age 40*. Ypsilanti, MI: HighScope Educational Research Foundation.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*(4), 317–329.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *Journal of Economic Perspectives*, *16*(3), 59–66.

Temple, J. A., & Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child–Parent Centers and related programs. *Economics of Education Review*, *26*, 126–144.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*(6), 681–694.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*(1), 122–154.