2016

A Benefit-Cost Analysis of the Tulsa Universal Pre-K Program

Timothy J. Bartik
W.E. Upjohn Institute, bartik@upjohn.org

Jonathan A. Belford
Child Trends

William T. Gormley
Georgetown University

Sara Anderson
West Virginia University

Upjohn Institute working paper ; 16-261

W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH
Researching the causes and consequences of unemployment

A Benefit-Cost Analysis of the Tulsa Universal Pre-K Program

Upjohn Institute Working Paper 16-261

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research
email: bartik@upjohn.org

Jonathan A. Belford
Child Trends
email: jbelford@childtrends.org

William T. Gormley, Jr.
Georgetown University
email: gormley@georgetown.edu

Sara Anderson
West Virginia University
email: sara.anderson@mail.wvu.edu

August 2016
Revised April 2017

ABSTRACT

In this paper, benefits and costs are estimated for a universal pre-K program, provided by Tulsa Public Schools. Benefits are derived from estimated effects of Tulsa pre-K on retention by grade 9. Retention effects are projected to dollar benefits from future earnings increases and crime reductions. Based on these estimates, Tulsa pre-K has benefits exceeding costs by about 2-to-1. This benefit cost ratio is far less than the benefit-cost ratios (ranging from 8-to-1 to 16-to-1) for more targeted and intensive pre-K programs from the 1970s and 80s, such as Perry Preschool and the Chicago Child-Parent Center (CPC) program. Comparing benefit-cost results from different studies suggests that our more modest estimates are due to two factors: 1) smaller percentage effects of pre-K on future earnings and crime in Tulsa than in Perry and CPC, and 2) smaller baseline crime rates in Tulsa than in the Perry and CPC comparison groups.

JEL Classification Codes: I28, I26, H43

Key Words: Universal pre-K, benefit-cost analysis, grade retention, crime education effects of education, increased earnings effects of education

A BENEFIT-COST ANALYSIS OF TULSA'S PRE-K PROGRAM

In recent years, many states and some local governments (e.g., New York City, Seattle, San Antonio) have significantly expanded pre-K funding. Are these short-term costs justified by long-term benefits? In this paper, we address this question by predicting the long-term benefits of the universal pre-K program in Tulsa, Oklahoma, through earnings increases and crime reductions, and comparing these benefits to the intervention's costs.

We focus on grade retention as a mechanism by which pre-K alters long-term outcomes. The connections are clear: as we show, Tulsa pre-K reduces grade retention, and grade retention reduces earnings and increases crime. Grade retention is an indicator for both "hard skills" (cognitive skills) and "soft skills" (social skills and other personality traits). If both are important, then grade retention is perhaps a more versatile indicator than most.

Our new Tulsa pre-K estimates are combined with national data for this benefit-cost analysis (BCA). We compare our results with previous BCAs of pre-K: Tulsa pre-K, projected from kindergarten test scores (Bartik, Gormley, and Adelstein 2012); the Perry Preschool program (Belfield, Nores, Barnett, and Schweinhart 2006; Heckman, Moon, Pinto, Savelyev, and Yavitz 2010); the Chicago Child-Parent Centers (CPC) program (Reynolds, Temple, White, Ou, and Robertson 2011); and a California universal pre-K proposal (Karoly and Bigelow 2005).

In our new Tulsa analysis, benefits exceed costs, but only modestly. Our benefit-cost ratios are much lower than more intensive and targeted pre-K programs, such as Perry and CPC. These lower ratios are due in part to Tulsa pre-K having lower percentage effects on adult outcomes compared to Perry and CPC. In addition, Perry and CPC targeted disadvantaged groups with high crime rates decades ago, allowing crime reduction benefits to be larger.

The next section of this paper reviews prior research. Our new estimates are then presented for Tulsa pre-K's effects on grade retention. Our methodologies are described for projecting retention effects on earnings increases and crime reductions, as well as for measuring program costs. Our overall BCA is then summarized, checked for robustness, and compared with previous studies. The conclusion considers implications for future research and policy.

LITERATURE REVIEW

Evidence on the Benefits of Preschool Participation

Many studies have examined the short-term effects of preschool, for "hothouse" programs run long ago (the Perry Preschool and Abecedarian programs), as well as for Head Start and state pre-K. These studies have converged on a firm conclusion—that high-quality preschool improves school readiness, sometimes dramatically (Gormley 2007; Camilli, Vargas, Ryan, and Barnett 2010; Yoshikawa et al. 2013).

Most pre-K research on school readiness has focused on disadvantaged students because many programs target this group. Some programs have shown solid test score gains (Frede, Jung, Barnett, Lamy, and Figueras 2007; Peisner-Feinberg, Schaaf, Hildebrandt, Pan, and Warnaar 2015; Puma, Bell, Cook, Heid, and Lopez 2005) while others have shown stunning gains (Ramey and Campbell 1984; Reynolds 2000; Weikart, Bond, and McNeil 1978). A few studies have examined universal pre-K, in Tulsa (Gormley et al. 2005), Boston (Weiland and Yoshikawa 2013), Georgia (Henry, Gordon, Henderson, and Ponder 2003), and Florida (Bassok and Miller 2014). The Tulsa and Boston programs enhance school readiness for middle-class as well as low-income students; the Georgia and Florida studies do not address this issue.

A few studies examine pre-K's longer-term effects and reach two conclusions: first, short-term effects on test scores diminish over the course of K-12, but do not disappear (Camilli et al. 2010; Duncan and Magnuson 2013); second, despite fading test score benefits, positive long-run benefits of pre-K are found for adult outcomes (e.g., educational attainment, earnings, crime). This medium- and long-run evidence is greatest for targeted programs. For universal programs, there is some medium-run evidence but no direct long-run evidence.

For targeted programs, medium-run studies in New Jersey (Barnett, Jung, Youn, and Frede 2013), Texas (Andrews, Jargowsky, and Kuhne 2012), and North Carolina (Dodge, Bai, Ladd, and Muschkin 2014; Peisner-Feinberg et al. 2015) find that pre-K participants had better test scores than nonparticipants in elementary school. An exception to these findings is a Tennessee study (Lipsey, Farran, and Hofer 2015), which finds no lasting test score benefits from pre-K after initial gains in kindergarten. Additionally, an experimental study of Head Start finds fade-out of short-term positive test score effects after kindergarten (Puma, Bell, Cook, and Heid 2010), although this conclusion has been challenged because many members of the control group also participated in preschool (Feller, Grindal, Miratrix, and Page 2016; Kline and Walters 2015). Gains are observed for the Head Start group when comparing them to children who receive home-based care (Feller et al. 2016). Overall, research on targeted programs suggests medium-run effects on test scores after kindergarten, but these effects are smaller than the effects at kindergarten (Camilli et al. 2010; Duncan and Magnuson 2013; WSIPP 2014).

For universal programs, evidence of medium-run test score effects is limited, because such programs are new and scarce. Studies of universal programs in Florida (Bassok and Miller 2014) and Tulsa (Hill, Gormley, and Adelstein 2015) do find such test score effects in elementary school.

For long-run effects on adult outcomes, some positive evidence is found for pre-K programs for the disadvantaged. The Perry study finds that treatment group members were more likely to graduate from high school, have a job, have higher earnings, and own a home, and were less likely to receive cash welfare and be arrested (Schweinhart, Montie, Xiang, Barnett, Belfield, and Nores 2005). The Abecedarian study finds that treatment group adults were more likely to graduate from high school, attend a four-year college, and have a job (Campbell, Pungello et al. 2012), but no crime or substance abuse effects were detected. The Chicago CPC study finds that program participants were more likely to graduate from high school, have a higher income, and have health insurance, and were less likely to engage in substance use and be arrested (Reynolds, Temple, Ou, Arteaga, and White 2011). However, these intense programs are not characteristic of today's pre-K programs, nor are the counterfactuals like today's due to the growing availability of social supports for four-year-old children.

Quasi-experimental evidence for the targeted and larger-scale Head Start program also suggests adult benefits, including increased earnings and reduced crime, although more for some subgroups than others (Deming 2009; Garces, Thomas, and Currie 2002).

For universal pre-K, no study to date provides direct evidence on adult outcomes because these programs are more recent and are harder to study using rigorous methodologies.

Finally, many studies find that pre-K reduces grade retention, a precedent for this paper, mostly focusing on targeted programs. For targeted programs, the Perry (Schweinhart et al. 2005), Abecedarian (Campbell and Ramey 1995; Campbell, Ramey, Pungello, Sparling, and Miller-Johnson 2002), and CPC studies (Reynolds et al. 2007; Reynolds, Temple, Robertson, and Mann 2001) find that pre-K reduced retention. Head Start is also found to reduce retention

(Deming 2009; Garces et al. 2002). More recent targeted studies also find that pre-K lowers retention in elementary school (Andrews et al. 2012; Barnett et al. 2013; Dodge et al. 2014).

For universal pre-K, only two studies address grade retention. In one study, universal pre-K reduces retention as of elementary school (Bassok and Miller 2014). In another study, universal pre-K reduces retention as of middle school (Phillips, Gormley, and Anderson 2016). The present paper extends Phillips, Gormley, and Anderson's (2016) study by using middle school retention effects to project long-term benefits.

In sum, evidence suggests a link between pre-K and favorable elementary, middle school, and adult outcomes. However, most studies have focused on targeted programs, not universal programs. Long-term effects on adult outcomes are not directly estimated for any universal program. This is a problem for BCA, because most benefits of pre-K are likely to be long-term. Because disadvantaged students have fewer resources than middle-class students, more modest benefits might be expected from universal programs. Our study helps fill this gap in the literature by linking universal pre-K to long-term effects via middle-term effects on grade retention, and then using these long-term projections to compare program benefits versus costs.

Previous Benefit-Cost Analyses of Pre-K Programs

A growing number of studies use pre-K's estimated effects to compare program benefits and costs.[1] Most pre-K benefits are due to earnings increases and crime reductions. BCAs of Perry Preschool from Belfield et al. (2006) and Heckman et al. (2010) show benefit-cost ratios of 16-to-1 and 8.5-to-1, with over 90 percent of benefits due to earnings increases or crime

[1] Karoly (2012) provides a review of benefit-cost methods and results for early childhood programs. Her conclusions about the high benefit-cost ratios for Perry and CPC are consistent with our discussion. All these studies implicitly take an efficiency perspective, with benefits and costs to everyone weighed equally. Yet we might also want to address distributional concerns. A pre-K program with efficiency benefits less than costs might still be chosen if it redistributes income to lower-income groups.

reductions. The BCA of CPC from Reynolds et al. (2011) finds a benefit-cost ratio of about 11-to-1, with over 75 percent of benefits due to earnings increases or crime reductions.

These high benefit-cost ratios for past programs may be of limited relevance to today's programs. Perry Preschool and CPC were well-funded and well-staffed and included home visits. Most of today's pre-K programs do not reach such standards (Mashburn et al. 2008). Also, the counterfactual for the comparison group has changed. In the 1960s and 1970s, child care subsidies for the disadvantaged were rare and Head Start enrollments were low. Today, child care subsidies are common and Head Start enrollments have climbed. In short, the educational exposure gap of treatment versus control children has narrowed over time.

Fewer BCAs have examined large-scale contemporary pre-K programs, including universal pre-K. The Washington State Institute for Public Policy (2014) reviewed research that included contemporary and earlier pre-K programs that were both targeted and universal, and projected average benefits versus costs. WSIPP (2014) concluded that, on average, benefits outweigh costs by 4-to-1 for state and local programs, and by 2.5-to-1 for Head Start.

Only two studies do a benefit-cost analysis of universal pre-K, and both studies do so without direct evidence on program effects on adult outcomes. Karoly and Bigelow (2005) estimate costs and benefits if California adopted high-quality universal pre-K and conclude that benefits would outweigh costs by 3-to-1. Karoly and Bigelow (2005) assume a universal program's benefits will be a fraction of CPC's benefits. Bartik, Gormley, and Adelstein (2012) estimate the earnings benefits of Tulsa pre-K and find benefit/cost ratios of over 3-to-1 for all income groups. These estimates use kindergarten test scores to predict future earnings, based on Chetty et al. (2011).

Overall, research suggests that current, larger-scale pre-K programs have benefits exceeding costs, but by much less than earlier, smaller programs of exceptionally high quality. The present study extends available research by estimating benefits and costs of universal pre-K in Tulsa, predicting benefits using medium-term effects on grade retention.

EFFECTS OF TULSA PRE-K PROGRAM ON GRADE RETENTION

The Tulsa Public Schools (TPS) pre-K program is universal, so all four-year-olds whose families reside in the district can enroll for free. Tulsa pre-K is high quality: every teacher has a BA degree, is early-childhood certified, and is on the same pay scale as other TPS teachers. Classroom observations suggest that, compared to 11 other states, Tulsa pre-K instructional quality is higher, with teachers performing better at providing students with feedback, language modeling, and higher-order concept development (Phillips, Gormley, and Lowenstein 2009). Benefits of Tulsa pre-K extend beyond academic achievement to social-emotional skills (Gormley et al. 2011).

Our BCA of Tulsa pre-K relies on its estimated effects on grade retention. These estimates are presented in more detail elsewhere (Gormley, Phillips, and Anderson 2016) but are summarized here.

Tulsa Sample and Model

Our sample starts with fall 2006 TPS kindergarten entrants (N = 4,033). The dependent variable was whether students were retained in grade one time as of 2015–2016, when students typically enter ninth grade. Grade retention was predicted using linear regression, in which the key independent variable was whether a student's enrollment in Tulsa pre-K in 2005–2006.

Students who attended Head Start were dropped from the sample, as they constituted their own treatment group. Of the original student sample, we could track 3,045 longitudinally through state records, with 1,283 having attended Tulsa pre-K and 1,410 not (others were in Head Start).

Observable variables were controlled for using propensity score weighting: independent variables were used to predict the probability of a student attending pre-K, and propensity score weights were used to weight control group observations so that the control group was similar to the treatment group in the probability of attending pre-K. Independent variables were derived from school administrative records, parent surveys, and the Census Bureau. They included gender, race, lunch status, mother's education and marital status, whether or not the child lived with the father, current school district, and neighborhood median income.

Using propensity score weights, we estimated effects of Tulsa pre-K on grade retention with linear probability regression models. We ran models for all pre-K students together and by gender, race/ethnicity (white/black/Hispanic), and free lunch status (free/reduced/paid). Models were also run by full- vs. half-day pre-K, and then by subgroup (Appendix A[2]). Estimates by both subgroup and full-day/half-day are omitted in the main paper due to limited sample sizes.

Pre-K and Grade Retention: Results

The results indicate that Tulsa pre-K reduced grade retention, both overall and for most subgroups (Table 1). Effects were stronger for more disadvantaged groups, such as blacks, Hispanics, and low-income groups. Effects were also stronger for males. However, effects only occurred for full-day pre-K (most enrollees) and not for half-day.

[2] The appendices are available online or from the authors.

Overall pre-K effects on retention are higher than effects of full-day pre-K or half-day pre-K. The reason for this pattern is that the propensity score weighting is done in three separate models: one model weights the control group to resemble those enrolled in any pre-K; two other models weight the control group to resemble those who attend full-day pre-K or half-day pre-K.

Which reweighting method is better? Propensity score reweighting only makes the control group resemble the treatment group on observable characteristics. In Tulsa, full-day pre-K is not available in all neighborhood schools. Parents can choose full-day pre-K by relocating or enrolling their child out-of-neighborhood and providing transportation. These choices may result in greater selection bias in the reweighting done separately for full- versus half-day pre-K. Parents who choose, or do not choose, full-day pre-K may differ in unobservables from the reweighted control group. Therefore, we prefer the estimates that reweight the controls to be comparable to the overall pre-K group, rather than the separate full- versus half-day groups. However, we report the full-day and half-day results for completeness and transparency.

ESTIMATING THE EFFECTS OF GRADE RETENTION ON EARNINGS AND CRIME

Our Tulsa data cannot be used to examine long-term earnings and crime outcomes, as the 2005–2006 Tulsa pre-K cohort was in ninth grade in 2015–2016. Therefore, we use our estimated pre-K effects on grade retention to predict adult earnings and crime outcomes.

Previous research suggests retention affects earnings and crime in part via effects on educational attainment (Jacob and Lefgren 2009). However, research is limited on the direct relationship between retention and these outcomes. Eide and Showalter (2001) find a negative relationship between retention and earnings, though Babcock and Bedard (2011) determine that retention during early grades (first and second) led to wage increases. Further, retained students

are arrested more than nonretained students, but McCoy and Reynolds (1999) find no relationship between grade retention and delinquency at age 14.

For this project, we provide new estimates of the link from grade retention to crime and earnings using a nationally representative study: the 1997 National Longitudinal Survey of Youth (NLSY97), which allows for direct estimates of how retention through eighth grade affects self-reported earnings and crime in young adulthood. We then project earnings and committed crimes over an individual's lifetime. Our estimated retention effects on earnings and crime could occur through educational attainment, as we do not hold attainment constant in estimating retention's effects, but could also occur through other mechanisms (e.g., behavior).

NLSY97 Data and Model

The NLSY97 is a longitudinal survey of 8,984 youth aged 12–17 in 1997, with an over-sample of black/Hispanic youth, and with follow-ups through 2013. Our analysis relates self-reported annual earnings and crimes to whether an individual had been retained in grade.

The dependent variables include, for each individual: 1) earnings for each age separately from 18–31; 2) the number of violent crimes, 1998–2013; and 3) the number of property crimes, 1998–2013.[3] Earnings are self-reported income from wages, salary, commissions, or tips from all jobs, before taxes. Crimes are self-reported arrests leading to police charges. Violent crimes include assault, rape, murder, and robbery. Property crimes include burglary and theft.

The independent variable of interest is whether a student has been retained at least once from first through eighth grade, similar to how the Tulsa data measures retention. The models include other control variables: gender, race/ethnicity, urban residence, birth cohort, English

[3] Because crime is a rarer event for individuals than having earnings, obtaining significant results for crime requires pooling data over multiple ages, whereas earnings determinants can be readily examined on an annual basis.

language learner status, free/reduced-price lunch status, single mother/two-parent family, and average highest grade completed of the respondent's parents.

An OLS model is used to predict the relationship between grade retention and earnings at each age, the number of violent crimes committed, and the number of property crimes committed. We ran models for youth overall, and by gender, race/ethnicity (white/black/ Hispanic), and the respondent's lunch status at baseline (free/reduced/paid). Overall sample sizes for earnings regressions ranged from 3,866 at age 18 to 1,363 at age 31; the sample size for the crime regressions was 4,324. Grade retention through grade eight in the overall sample averaged 15 percent. Sample sizes were reduced due to missing data, a decline in response rate to 80 percent by 2013, and later ages having data only for older birth cohorts.[4]

Effects of Grade Retention on Earnings and Crime: Summary of Results

As expected, grade retention was associated with lower earnings for each age from 18 to 31. Negative effects for most ages/groups were statistically significant; for the 126 estimated retention effects on earnings (9 groups/14 ages), 104 were significant at the 5 percent confidence level. Grade retention was also associated with higher crime, although crime effects were often insignificant. For the overall sample, retention was significantly associated (5 percent) with increased violent and property crime. For violent crime, retention's positive association was significant for white youth and marginally significant (10 percent) for the male and paid lunch groups. For property crime, retention's positive association was only marginally significant for males. Tables 2 and 3 summarize the results. Appendix B provides additional evidence.

[4] Appendix B compares means between the full sample and the regression samples at ages 18, 25, and 31. Although the regression and full samples have some statistically significant differences, the sizes of these differences are small.

Figures 1 and 2 interpret the earnings results. These figures convert dollar effects into percentage effects at mean earnings of each group/age. For the overall sample, Figure 1 shows percentage earnings effects by age. Grade retention reduces earnings by 20–30 percent, with no obvious trend. For each group, Figure 2 shows the average percentage effect (over all ages, 18–31). For each group, retention reduces earnings by 15–35 percent. Percentage earnings effects were greater for disadvantaged groups, such as racial minorities and lower-income groups.

Figure 3 interprets the crime results. Estimated effects are shown as percentage effects at the mean of each group's reported crimes. For the overall sample, retention increases self-reported arrests by 60–70 percent. Percentage effects of retention on crime are larger for more advantaged groups, such as whites and the paid lunch group. Percentage effects on crime are similar for males and females, although this corresponds to a larger absolute effect for males.

These large retention effects on earnings and crime reflect that we allow for retention effects via any post-eighth-grade outcomes. For example, a portion of retention's "effects" in our estimation occur due to retention's correlation with lower educational attainment.

Transferability of NLSY97 and Tulsa Pre-K Estimates

While these regressions cannot prove that grade retention causally reduces earnings and increases crime, these estimates are plausible. Holding constant other determinants of earnings and crime, retention seems to predict adverse outcomes in adulthood.

Are these estimates transferable to changes in grade retention due to Tulsa pre-K? Retention in both cases is measured as of eighth grade; it may predict different adult outcomes if measured in earlier grades versus later grades. Although head-to-head comparisons of retention at different grades are lacking, evidence suggests that retention in elementary school may be

more harmful than retention in middle school (Andrew 2014; Jacob and Lefgren 2009). Therefore, measuring retention for the same grades may make the estimates more transferable.

Even with retention measured at the same grade, the NLSY97 estimates might overstate or understate the effects of Tulsa pre-K on earnings and crime. For example, students retained in grade often have major learning or behavioral problems that would dramatically reduce earnings and increase crime. Tulsa pre-K might only reduce retention for students with more modest problems, which might imply lesser effects on adult outcomes.

On the other hand, NLSY97 includes retention data from students in many U.S. states. Some variation in retention of NLSY97 youth might be due to differences in state/district policies rather than the retained individuals having worse problems. In contrast, differences in retention induced by Tulsa pre-K probably reflect some improvement in learning and behavior, as retention policy will be more similar within one state than across states. Differences in retention due to different state policies might have lesser effects on adult outcomes than differences in retention due to pre-K in one state.

Grade retention policy in the U.S. changed between the NLSY97 cohorts (age 12–17 in 1997) and our Tulsa cohort (ages 12–17 in 2013–2018). Data from the Current Population Survey shows that annual retention rates from kindergarten to eighth grade declined from 3.1 percent in 1994 to 2.3 percent in 2015 (Table 225.90, *Digest of Education Statistics*). If declining retention rates are due to policy, then it takes worse behavior/learning to be retained in later cohorts. But it is unclear whether this policy change will result in larger or smaller effects of pre-K via retention effects. The average behavior/learning difference between a student retained or not retained because of pre-K could be either larger or smaller today than in the 1990s.

The large retention effects in NLSY97 have indirect as well as direct effects on adult outcomes. Our NLSY97 estimates do not hold constant any outcomes after eighth grade, such as educational attainment. Therefore, our large retention effects include effects due to lower educational attainment. This is a feature and not a drawback of our approach. We use retention to proxy for subsequent outcomes, and thereby to predict adult earnings and crime.

Ultimately, the validity of using NLSY97 retention effects to predict long-run effects of Tulsa pre-K can only be known when adult follow-up data for the Tulsa Pre-K cohort are available. At present, our only test of validity is to compare these retention-derived predictions of pre-K benefits with predictions based on eighth grade test scores (Table 6 and Appendix C).

PROGRAM BENEFITS

The NLSY97 results only show earnings and crime effects of grade retention through early adulthood. For benefit-cost analysis, effects must be projected over a lifetime. This section discusses these projections. In addition, the crime reduction benefits require one additional step: assigning crime reductions a dollar value.

Predicting Earnings Benefits

To project NLSY97 retention effects on young adult earnings to a lifetime, we first determine baseline lifetime earnings for each group, using data from the American Community Survey (ACS), 2009–2013, for persons in the Tulsa metro area. Using Tulsa ACS data, age-earnings profiles are calculated for the overall sample and for the gender and racial groups.[5] Data

[5] The ACS data are already adjusted to 2013 Tulsa prices. We adjust to national prices using regional price parities from the U.S. Bureau of Economic Analysis (2016).

are adjusted to 2013 national prices to allow comparisons with previous pre-K benefit-cost studies, which we convert to the same standard. Mean earnings by single year of age for ages 18–79 are calculated for the overall sample, separately for men and women, and for three racial groups: white non-Hispanic, black non-Hispanic, and Hispanic.

ACS data do not show how age-earnings profiles vary by family income of an individual when growing up. Individuals who grew up lower-income would tend to have lower earnings. To project how future earnings vary with family income background, ACS data on Tulsa earnings are combined with NLSY97 data on earnings and past family income. In NLSY97, we calculate for individuals at ages 12–17 whether their families' incomes qualify them for free or reduced-price lunches, or no lunch subsidy. For these three family income groups, and the overall sample, we calculate average earnings for each age from 18 to 31. We then calculate the ratio of each group's earnings to the overall average, for each age.

We assume these NLSY97-derived ratios show how earnings vary by an individual's family income background in the ACS. Therefore, we take average overall earnings in the ACS, for each year of age from 18 to 31 and multiply these by the NLSY-derived ratios to predict earnings for the different income background groups. For ages beyond 31, we project earnings by multiplying the ACS overall earnings for each age from 32 to 79 by the average NLSY97 ratio of ages 25–31 for each income background group (65/85/117 percent for free/reduced/paid lunch). Averaging over ages 25–31 reduces estimate volatility; starting with age 25 reflects that more individuals have completed education at 25 than earlier ages.

These age-earnings profiles only show how earnings vary with age in 2009–2013. However, our BCA needs to adjust for mortality since age 4 and secular future earnings increases. To adjust for mortality, we use U.S. Life Tables from 2011 (Arias 2015). We calculate

ratios of the expected number of persons alive at each age to the persons alive at age 4.[6] We adopt the midrange assumption of the Social Security Trustees that long-run growth in U.S. real wages will be 1.17 percent per year over the next 75 years (U.S. Social Security Administration 2015).

To compare earnings benefits with pre-K costs at age 4, we discount future earnings back to age 4. We use a 3 percent annual real discount rate, which is commonly used in BCAs.

Given recent trends, earnings growth of 1.17 percent could be questioned. But, what matters for the present value of earnings is the difference between the discount rate and earnings growth. It is hard to justify a discount rate as great as 3 percent if earnings will not increase at least 1 percent annually. Discounting is justified because future citizens will be wealthier, reducing the value of future income with stagnant earnings. As a result, the discount rate should be lower than 3 percent, with little net effect on earnings' present value.[7]

To calculate how retention affects lifetime earnings, the age-earnings profiles of each group are combined with the percentage effects of retention on earnings, from the previous section. For ages 18–31, we multiply the NLSY97-percentage effect by the adjusted ACS earnings for that age/group to get the retention effect for that age/group. For ages 32–79, we multiply the adjusted ACS earnings for that age/group by that group's NLSY97-derived average percentage effects for ages 25–31. When discounted and summed over all ages, this calculation gives the present value reduction in lifetime earnings from grade retention.

This present value of the earnings reduction from retention for each group is then multiplied by the estimated pre-K effects on retention to give the estimated pre-K effects on the

[6] Life Tables are unavailable for family background so we use black data for the free/reduced-price lunch groups, and white data for the full price lunch group.

[7] See Bartik (2011) for further discussion.

present value of earnings. This gives us the earnings benefits of Tulsa pre-K participation mediated through its effects on grade retention, overall and by group.

Predicting Crime Benefits

Previous sections showed negative effects of pre-K on retention and positive effects of retention on self-reported crime in young adulthood. To translate these estimates into dollar benefits of crime reduction, we need two more pieces of information: baseline actual crime (likely greater than self-reports) at different ages and the social costs per crime committed.

To project baseline Tulsa crime rates, we use data from the 2012 FBI Uniform Crime Reports (UCR), which reports data on arrests and total crimes reported, broken down by type of crime: murder, rape, assault, robbery, theft, and burglary. The first four are "violent crimes" and the last two are "property crimes." Arrests are broken down by age and gender and are used to allocate crime counts by age and gender. Crime counts are combined with census population data to calculate crime rates by age, gender, and type of crime.[8] As expected, crime rates are highest for persons in their late teens and early twenties, and are much higher for males than females.

To derive crime rates by race and age, overall crime counts by age are multiplied by relative Tulsa crime rates by race. Relative crime rates for whites and blacks are calculated from the Tulsa UCR data on arrests and Census population data.[9] The UCR does not report arrests by Hispanic status. Hispanic crime rates are assumed to be the same as the overall population.[10]

To calculate crime rates by age for the different family income background groups, the overall age-crime profiles, calculated above, are multiplied by relative arrest rates by family

[8] Because some data are reported by age ranges, interpolation is required.

[9] Arrest data by race are used to allocate crimes by race and age. Census individuals were counted as white if indicating no other race and as black even if indicating multiple races.

[10] There is little data about Hispanic crime. First-generation immigrants have lower crime, but crime rates converge to U.S. averages for subsequent generations. See Brame et al. (2014) and Morin (2013).

income background group, derived from the NLSY97. Using NLSY97 data, relative arrest rates are calculated from 1998 to 2013, for violent and property crimes separately, broken down by the individual's family income at ages 12–17. Crime rates are 50–75 percent higher for individuals from a free lunch background than for the average individual, and about half as high for individuals from a paid lunch background than for the average individual; individuals from a reduced-price lunch background are close to the overall average.

For the benefit-cost calculations, these crime rate figures need to be adjusted for mortality since age 4. We use the same mortality assumptions as was done for earnings.

Social costs per each type of crime are chosen to be in the middle of the research literature. We rely on McCollister, French, and Fang (2010, Table 1), who report crime cost estimates from six other studies and provide their own estimates. From these seven studies, we choose the median social cost for each type of crime.

To calculate present values of crime costs by group, we first multiply the expected number of each crime type per person for each group and single year of age, adjusted for mortality, by median social cost of that crime. Costs are discounted by 3 percent back to age 4.

For each group, present values of crime costs by type are summed into the broader violent crime and property crime categories. For violent crime and property crime separately and for each group, percentage effects of Tulsa pre-K are estimated by multiplying effects of pre-K on retention by percentage effects of retention on crime. The present value of expected future crimes in each category for a typical 4-year-old in each group is then multiplied by the percentage effects of crime reduction for that category/group. We then sum the violent and property crime categories. This yields estimated benefits of reduced crime from Tulsa pre-K participation, mediated through grade retention, for the overall sample and subgroups.

Note that these procedures base the benefits of crime reduction on crimes reported to the police. This may understate true crime, but less severely than for self-reported arrests in the NLSY.[11] The arrests data is used to allocate the reported crime by age and group, and the NLSY97 estimates are used to calculate the percentage reductions in crime. The benefits calculations are adjusted to the higher police-reported crime standard.

Other Benefits

Our benefit measures for Tulsa pre-K are conservative because they only include increased earnings and reduced crime but not other benefits. If pre-K lowers retention, it lowers public costs of providing more years of education to students who would have been retained. Pre-K may also provide benefits by: lowering special education costs; increasing earnings of parents; lowering health care costs of former pre-K participants; reducing welfare participation of former pre-K participants; and savings in reduced private pre-K and child care costs.

These other benefits in previous pre-K studies are of modest size, relative to the benefits of higher earnings and lower crime. In existing studies, other benefits are less than one-quarter of total benefits. However, a more complete BCA might find a larger role for other benefits.

PROGRAM COSTS

An ideal measure of the costs of pre-K would be contemporaneous (gathered just after costs were incurred), inclusive (encompassing all costs), and authoritative (supplied by knowledgeable observers). Personnel costs, capital costs, and in-kind costs would all be

[11] Farrington (1992) provides a review of difficulties of estimating lifetime crime.

included. Differential costs for different categories of students would be considered, and cost differentials between full-day and half-day programs would be factored into the analysis.

In practice, it is difficult to find a perfect data source. Nevertheless, we do have access to three credible sources. Based on them, we derive three estimates of Tulsa pre-K costs in 2013 U.S. dollars, which allows comparison with previous pre-K studies (Table 4).[12]

First, a report by the Institute for Women's Policy Research (Gault et al. 2008) estimated the costs of high-quality pre-K. Quality is defined in multiple ways, based on teacher qualifications, classroom size, etc. Thus, one can select the option best resembling the pre-K program being studied. For our purposes, that would be one with 20 students per class, led by a teacher with a bachelor's degree and paid public school wages. The Institute for Women's Policy Research report includes both personnel and capital costs. It assumes a school year of 185 days and distinguishes between half-day (three hours) and full-day (six hours) programs.

Second, a report by the National Institute for Early Education Research (NIEER 2007) has estimated the costs of state-funded pre-K for each state with a program, including Oklahoma. These annual reports begin in 2003. Given our focus on the TPS 2005–2006 pre-K program, we utilize the 2007 report, which provides 2005–2006 expenditure data, gathered through a survey of state pre-K administrators, and which estimates average pre-K costs in 2006 Oklahoma dollars. It is difficult to determine from NIEER's report whether pre-K costs include capital costs. Also, there is no separate calculation for half-day and full-day programs. This omission can be corrected by assuming a 1.86 cost differential between full-day and half-day programs, based on the state aid formula, which assigned a weight of 1.3 for full-day students and 0.7 for

[12] Price adjustments use data from the U.S. Bureau of Economic Analysis. Adjustments to 2013 use the price deflator for GDP personal consumption expenditures. Adjustments from Tulsa/Oklahoma to U.S. prices use regional price parities.

half-day students in 2005–2006. We assume the mix of full-day and half-day pre-K students is the same for Oklahoma as for Tulsa.

Third, data from the U.S. Census Bureau, combined with information from the TPS Treasurer, Joe Stoeppelwerth, yielded cost estimates that were Tulsa-specific. We began with Census Bureau data for TPS on total personnel costs (teachers and others) for all grades in the 2005–2006 school year (U.S. Census Bureau 2006). Studies suggest that personnel costs are 80 percent of the average school district's budget (Cavanagh 2011; American Association of School Administrators, n.d.). Therefore, we estimated total TPS school district costs by dividing total personnel costs by 0.80. Next, we used data from TPS to estimate the ratio of pre-K expenses to total district expenses. It is unclear how the cost of pre-K compares to that of K-12 grades. While pre-K requires lower student/teacher ratios than other grades, which drives up costs, pre-K also typically has less-experienced teachers, which drives down costs. In TPS, the ratio of pre-K expenses to total school district expenses was 0.033, which we used to estimate pre-K expenses. Based on the state aid formula, we assume a full-day program costs 1.86 times as much as a half-day program. Beyond the monetary costs of pre-K to TPS, there may be some in-kind resources from parents, but we believe such in-kind costs to be slight, given that Tulsa pre-K is run by a public school district with a high percentage of lower-income parents.

When calculating benefit/cost ratios, we prefer the Tulsa numbers because they come more directly from the school district and yield more conservative benefit/cost estimates. Clearly, however, the numbers, from three separate sources, converge to a striking degree.

These half-day and full-day cost figures are directly used in our BCA of half-day and full-day pre-K. For the overall sample, and for the different subgroups, the cost figure per child is adjusted based on the observed mix of half-day versus full-day enrollment.

Our cost estimates assume zero costs in the comparison group. This is appropriate in terms of public pre-K costs, as the comparison group excludes Head Start enrollees. However, there may be private pre-K costs in the comparison group, which would be greater than private pre-K costs in the Tulsa pre-K group. These cost savings on private pre-K are another benefit of publicly provided pre-K, but one which is not measured in this BCA.

BENEFIT-COST ANALYSIS RESULTS

Main Results

Combining the benefit and cost results from the previous sections gives us BCA calculations, shown in Table 5. Tulsa pre-K for most groups appears to pass a benefit-cost test, but only modestly. The overall program benefit-cost ratio is 2.10.

Earnings benefits are much greater than crime reduction benefits. For example, in the overall results, the ratio of earnings benefits to crime reduction benefits is over 4-to-1.

The results in Table 5 provide mixed evidence on whether pre-K's net benefits are larger for the disadvantaged. Benefit-cost ratios and net benefits tend to be higher for pre-K participants from lower income groups compared to the highest income group. They are also relatively high for Hispanics. However, the ratios are somewhat higher for the average white pre-K participant than for the average black participant. Net benefits are much greater for males than for females.

Results are more favorable for full-day than for half-day pre-K. Net benefits of half-day pre-K are negative. Note that overall results are not an average of the full-day and half-day results. As discussed previously, this occurs because of differences in the matching algorithm when matches are done separately for full-day versus half-day, versus all pre-K together.

Robustness Checks

These benefit-cost results rely on particular estimates, assumptions, and methodologies. How sensitive are these results to alternatives?

Table 6 summarizes how plausible alternatives affect benefit-cost ratios, including:

- Different point estimates. The baseline results are derived from estimates of Tulsa pre-K effects on retention. These estimates are uncertain.[13] What is the confidence interval for our benefit-cost results, as implied by the estimates' standard errors?
- Earnings growth assumptions. The baseline results assume that future real earnings will grow at about 1 percent annually. What if future real earnings do not grow?
- Social cost of crime assumptions. The baseline results use the median social cost of crime, from seven studies in McCollister, French, and Fang (2010). How do results differ if we use the lowest/highest costs of crime in these seven studies?
- Projecting earnings benefits from test scores, not grade retention. The baseline results predict pre-K's effects on adult earnings using retention by eighth grade. What if instead we use pre-K's estimated effects on eighth grade test scores? Test scores can be used to predict earnings using several studies (e.g., Chetty et al. 2011).

Appendix C provides detail on these robustness checks. As shown in Table 6, the robustness checks do not dramatically change results for the overall sample. The overall confidence interval suggests that program benefits exceed costs, even with extreme parameter estimates. Benefit/cost ratios are modestly reduced with lower earnings growth, or lower crime

[13] There also is uncertainty in: NLSY estimates; ACS earnings estimates; crime/arrest estimates; life table estimates; estimates of crime's social costs; secular earnings growth. Addressing all uncertainty simultaneously would be challenging; we address a few uncertainties in robustness checks.

costs, and modestly increased with higher crime costs. The overall benefit-cost ratio is similar if test scores are used instead of grade retention to predict future earnings benefits.

Results are more sensitive for different groups. There is more uncertainty in group estimates, so confidence intervals are wider. Estimated effects of Tulsa pre-K on eighth grade tests are noisier than retention effects, so estimates based on tests vary widely.

Comparing Overall Results to Previous Studies

Comparisons with previous studies are in Table 7. Our current results are compared with an earlier study of Tulsa pre-K, which projected benefits based on kindergarten test scores (Bartik, Gormley, and Adelstein 2012). These Tulsa results are compared with two BCAs of Perry Preschool, Belfield et al. (2006) and Heckman et al. (2010); and a BCA of Chicago CPC (Reynolds, Temple, White et al. 2011). The Perry studies and the CPC study are based on data on adult outcomes but projected to the entire life-cycle. Results are also compared with a BCA for a proposed California universal pre-K program (Karoly and Bigelow 2005), with these estimates based on scaling back the CPC estimates for the presumed lower benefits of a universal program.

In Table 7, we focus on results for each study's overall sample. Results for subgroups are in Appendix D. Because this current study only considers earnings benefits and crime reduction benefits, we report each study's benefit-cost ratio with and without other benefits.

The Perry and CPC studies have much higher benefit-cost ratios. Our current study has an overall benefit-cost ratio for Tulsa of 2.10; the prior Tulsa study had a benefit-cost ratio of 3.20. In contrast, the Perry Preschool benefit-cost ratio is over 8-to-1 in Heckman et al. (2010) and over 15-to-1 in Belfield et al. (2006), even when only counting earnings increases and crime reduction as benefits. For Chicago CPC, the comparable benefit-cost ratio is over 8-to-1.

In the California study of universal pre-K, overall net benefits and the benefit-cost ratio are quite similar to the current study. Benefits and costs are somewhat lower in the California study because it assumes that benefits can be achieved with lower-cost half-day pre-K, whereas our current study suggests that high benefits require full-day pre-K.

In sum, our current study suggests real but modest support for the universal Tulsa pre-K program as a good investment. These estimates are similar to the proposed California universal pre-K program. In contrast, the prior studies of Perry and CPC suggest extraordinarily high rates of return, but for past smaller-scale programs highly targeted on the disadvantaged.

Differences in benefits between the current study and the Perry and CPC studies are in part due to earnings benefits. Earnings benefits in CPC are twice as great as the current study. For Perry, overall earnings benefits are five or six times as great as the current study.

However, most differences between the current study and the Perry and CPC studies are due to crime reduction benefits. CPC crime reduction benefits are over 15 times greater than the current study, and Perry crime reduction benefits are over 25 times as great (Heckman et al. 2010) or over 70 times as great (Belfield et al. 2006).

What is behind these large differences? Appendix E explores this subject in more detail and concludes the following: compared to the current Tulsa study, the estimated percentage effects of the Perry and CPC programs on increasing earnings and reducing crime, are much higher, from 4 to 10 times as great. Also, baseline crime rates in the Perry and CPC studies are much greater, by 5 to 10 times. Higher baseline crime will increase a program's benefits holding constant its percentage effects on crime reduction.

CONCLUSION

This study of Tulsa's universal pre-K program finds that the program's benefits exceed costs. This finding is conservative because we omit some benefits. The policy implication is that pre-K programs that are universal and high quality will provide net benefits to society.

However, the benefit-cost evidence for Tulsa's universal pre-K is not as favorable as for small, intensive, and highly targeted programs such as Perry and CPC that were implemented decades ago. One cannot promise policymakers that large, less-intensive universal programs will deliver the extraordinary results of Perry-style programs. However, extraordinary benefits are not needed for a program to be worthwhile.

One key to the more modest benefits of today's universal pre-K, compared to past targeted programs, is more modest baseline crime rates. In designing pre-K, the baseline crime rate of the targeted group can have large implications for benefits. Lower baseline crime rates in today's universal pre-K, compared to past targeted programs, may reflect both lower expected crime among the universal program's participants, as well as secular declines in crime.

Although our findings on Tulsa's universal pre-K are more modest than for early iconic programs, the Tulsa program has the advantage of being a large-scale contemporary program that reaches a large and diverse cross-section of children. Furthermore, the costs of Tulsa pre-K could be financed by a typical public school district. Thus, our research has higher external validity than other studies. However, we should caution that we have examined only two pathways for pre-K impacts: from pre-K to grade retention to earnings, and from pre-K to retention to crime. If pre-K produces benefits through other pathways or influences other outcomes, then our estimated program impact is a lower boundary for actual impact. With longer-run follow-up of

pre-K participants into adulthood, future research may be able to provide more direct and accurate estimates of pre-K's adult benefits.

REFERENCES

American Association of School Administrators. (n.d.). School budgets 101. https://www.aasa.org/uploadedFils/Policy_and_Advocacy/files/SchoolBudgetBriefFINAL.pdf (Accessed August 17, 2016.)

Andrew, M. (2014). Primary-grade retention? A study of cumulative advantage in the educational career. *Social Forces, 93*, 653–685.

Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). *The effects of Texas's targeted pre-kindergarten program on academic performance.* (NBER Working Paper No. 18598.) Cambridge, MA: National Bureau of Economic Research.

Arias, E. (2015). United States life tables, 2011. *National Vital Statistics Reports*, *64*(11), 1–62.

Babcock, P., & Bedard, K. (2011). The wages of failure: New evidence on school retention and long-run outcomes. *Education and Finance Policy, 6*(3), 293–322.

Barnett, W. S., Jung, K., Youn, M., & Frede, E. C. (2013). *Abbott preschool program longitudinal effects study: Fifth grade follow-up*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

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

Bartik, T. J., Gormley, W. E., & Adelstein, S. (2012). Earnings benefits of Tulsa's pre-K program for different income groups. *Economics of Education Review, 31*(6), 1143–1161.

Bassok, D., & Miller, L. C. (2014). *The effects of Florida's voluntary pre-kindergarten program on children's early elementary school outcomes*. Paper presented at annual meeting of the Association for Public Policy Analysis & Management, Albuquerque, NM.

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. (2006). The High/Scope Perry preschool program: Cost-benefit analysis using data from the age–40 followup. *Journal of Human Resources, 41*(1), 162–190.

Brame, R., Bushway, S. D., Paternoster, R., & Turner. M. G. (2014). Demographic patterns of cumulative arrest prevalence by ages 18 and 23. *Crime & Delinquency, 60*(3), 471–486.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579–620.

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., Barbarin, O. A., Sparling, J. J., & Ramey, C. T. (2012). Adult outcomes as a function of an early childhood education program: An Abecedarian project follow-up. *Developmental Psychology, 48*(4), 1033–1043.

Campbell, F. A., & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32*(4), 743–772.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6*(1), 42–57.

Cavanagh, S. (2011, January 13). Personnel costs prove tough to contain. Retrieved from http://www.edweek.org/ew/articles/2011/01/13/16personnel.h30.html

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.

Deming, D. J. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111–134.

Dodge, K. A., Bai, Y., Ladd, H., & Muschkin, C. (2014). *Impact of statewide early childhood programs and policies on educational outcomes through the end of elementary school.* Paper presented at annual meeting of Association and Public Policy Analysis & Management, Albuquerque, NM.

Duncan, G., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*(2): 109–132.

Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review, 20*(6), 563–576.

Feller, A., Grindal, T., Miratrix, L. W., & Page, L. C. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care-type settings. http://dx.doi.org/10.2139/ssrn.2534811 (accessed August 17, 2016).

Frede, E., Jung, K., Barnett, W. S., Lamy, C., & Figueras, A. (2007). *The Abbott preschool program longitudinal effects study (APPLES).* New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *American Economic Review, 92*(4), 999–1012.

Gault, B., Mitchell, A., Williams, E., Dey, J., & Sorokina, O. (2008). Meaningful investments in pre-K: Estimating the per-child costs of quality programs. Institute for Women's Policy, Washington, DC.

Gormley, W. T. (2007). Early childhood care and education: Lessons and puzzles. *Journal of Policy Analysis and Management, 26*(3), 633–671.

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

Gormley, W. T., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development, 82*(6), 2095–2109.

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

Henry, G. T., Gordon, C. S., Henderson, L. W., & Ponder, B. D. (2003). *Georgia pre-K longitudinal study: Final report 1996–2001.* Atlanta, GA: Andrew Young School, Georgia State University.

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

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1*(3), 33–58.

Karoly, L. (2012). Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis, 3*(1): 1–45.

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

Kline, P., & Walters, C. (2015). *Evaluating public programs with close substitutes: The case of Head Start*. (NBER Working Paper No. 21658). Cambridge, MA: National Bureau of Economic Research.

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749.

McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence, 108*, 98–109.

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*(3), 273–298.

Morin, R. (2013, October 15). Crime rises among second-generation immigrants as they assimilate. *Factank, Pew Research Center*. http://www.pewresearch.org/fact-tank/2013/10/15/crime-rises-among-second-generation-immigrants-as-they-assimilate/ (accessed August 17, 2016).

National Institute for Early Education Research. (2006). *The State of Preschool 2006: State Preschool Yearbook*. http://nieer.org/state-preschool-yearbooks/state-preschool-2006#contents (accessed August 17, 2016).

Peisner-Feinberg, E. S., Schaaf, J. M., Hildebrandt, L. M., Pan, Y., & Warnaar, B. L. (2015). *Children's kindergarten outcomes and program quality in the North Carolina pre-kindergarten program: 2013–2014 statewide evaluation*. Chapel Hill, NC: University of North Carolina, Frank Porter Graham Child Development Institute.

Phillips, D., Gormley, W., and Anderson, S. (2016). The effects of Tulsa's CAP Head Start program on middle-school academic outcomes and progress. *Developmental Psychology, 52*(8), 1247–1261.

Phillips, D. A., Gormley, W. T., & 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.

Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start impact study: Final report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Puma, M., Bell, S., Cook, R., Heid, C., & Lopez, M. (2005). *Head Start impact study: First year findings*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian project. *American Journal of Mental Deficiency, 88*(5), 515–523.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago child-parent centers.* Lincoln, NE: University of Nebraska Press.

Reynolds, A. J., Temple, J. A., Ou, S. R., Arteaga, I. A., & White, B. A. (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., Ou, S. R., Robertson, D. L., Mersky, J. P., Topitzes, J. W., & Niles, M. D. (2007). Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. *Archives of Pediatric and Adolescent Medicine, 161*(8), 730–739.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational attainment and juvenile arrest. *Journal of the American Medical Association, 285*(18), 2339–2346.

Reynolds, A. J., Temple, J. A., White, B. A., 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.

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. *Monographs of the HighScope Educational Research Foundation* 14. Ypsilanti, MI: High/Scope Press.

U.S. Bureau of the Census (2006). *Public School System Finances: Historical Data, 2006.* https://www.census.gov/govs/school/historical_data_2006.html (accessed March 29, 2017).

U.S. Bureau of Economic Analysis. (2016). *Regional price parities database.* http://www.bea.gov/regional/index.htm (accessed June 9, 2016).

U.S. Social Security Administration. (2015). *The long-range economic assumptions for the 2015 trustees report*. Washington, DC: U.S. Social Security Administration.

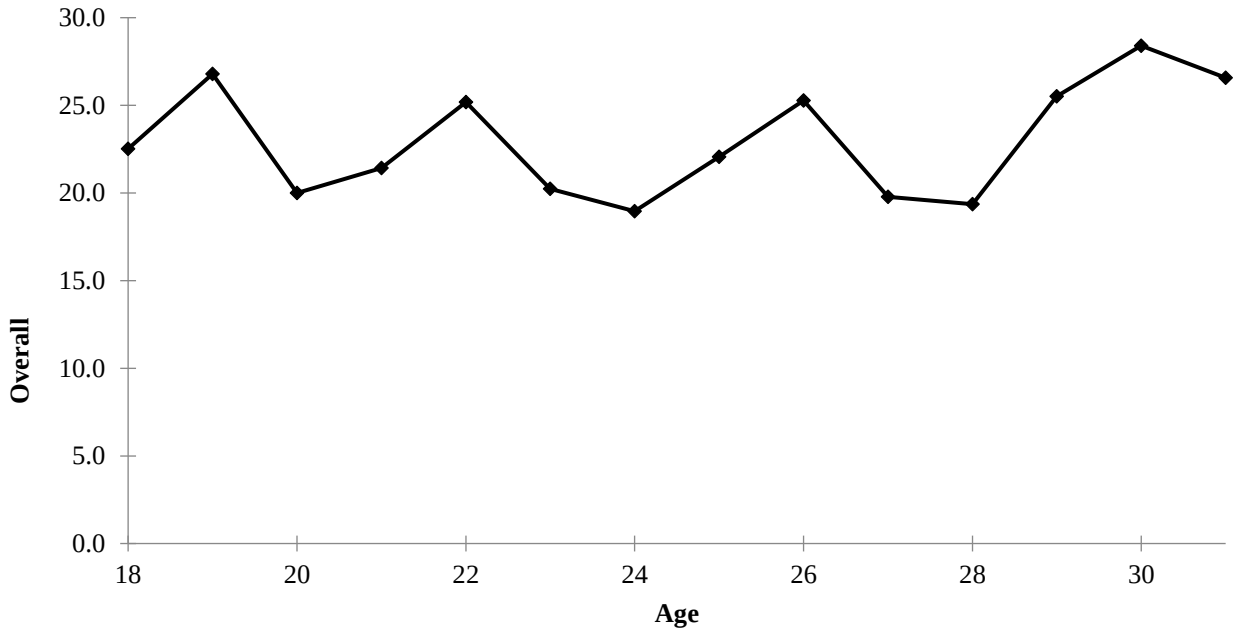
Washington State Institute for Public Policy. (2014). *Early childhood education for low-income students: A review of the evidence and benefit-cost analysis*. Olympia, WA: Washington State Institute for Public Policy.

Weikart, D. P., Bond, J. T., & McNeil, J. T. (1978). *The Ypsilanti Perry preschool project: Preschool years and longitudinal results through fourth grade*. Ypsilanti, MI: High/Scope Press.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112–2130.

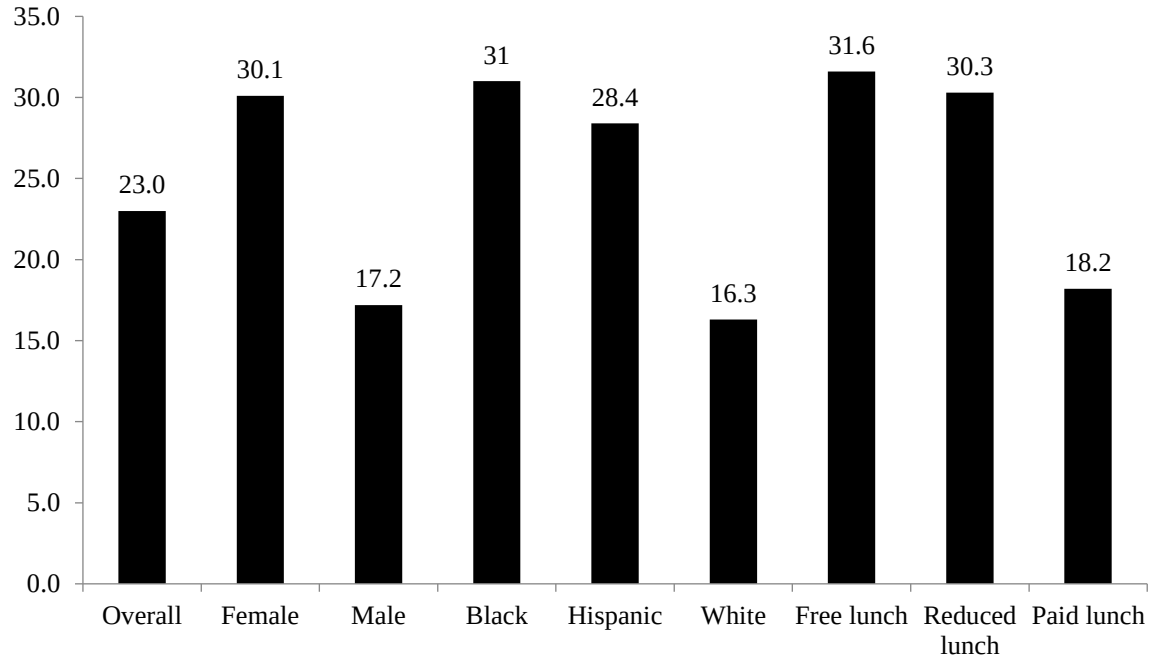
Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. H. (2013). *Investing in our future: The evidence base on preschool education*. New York: Foundation for Child Development.

Figure 1 Earnings Reductions Due to Grade Retention (%)



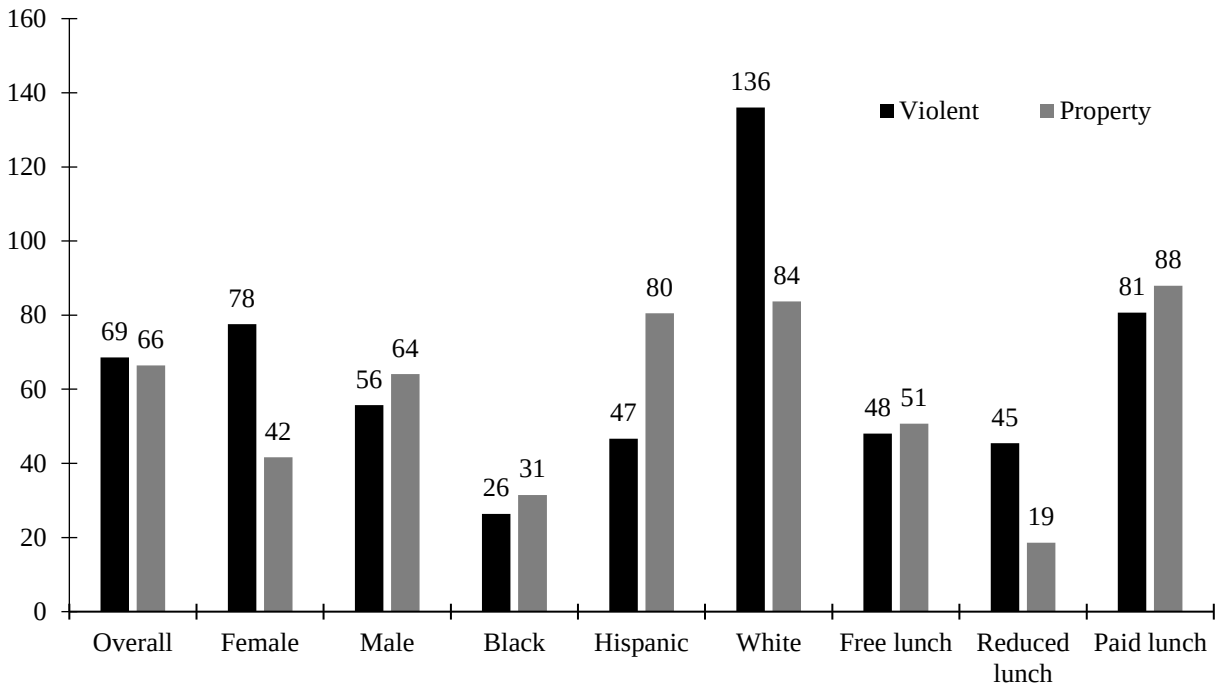
NOTE: Derived from NLSY97 estimates, as described in text, and more fully presented in Appendix B. The percent results are for the overall NLSY97 sample. They are derived by dividing the dollar reduction estimates for each age by estimated mean earnings by age. Estimates control for gender, race/ethnicity, urbanicity, birth cohort, ELL status, free/reduced-price lunch status, single mother/two-parent family, and parental education. All estimates are statistically significant at 1 percent level.
SOURCE: Authors' estimates.

Figure 2 Average Percent Reduction in Earnings Due to Grade Retention, Ages 18–31



NOTE: Derived from estimates described in text, based on NLSY97. Estimates based on taking reduction in dollar earnings for single years of age and group, and dividing by mean earnings for that age and group. Estimates control for gender, race/ethnicity, urbanicity, birth cohort, ELL status, free/reduced-price lunch status, single mother/two-parent family, and parental education. Average is simple average of those percent reductions over ages 18-31.
SOURCE: Authors' estimates.

Figure 3 Percent Increase in Crime Due to Grade Retention, by Group and Type of Crime



NOTE: This figure is derived from NLSY97 estimates, as described in text. Percent increase in crime due to grade retention is the estimated effect on arrests from 1998 to 2013, divided by mean arrests for that group over that time period. Estimates control for gender, race/ethnicity, urbanicity, birth cohort, ELL status, free/reduced-price lunch status, single mother/two-parent family, and parental education.
SOURCE: Authors' estimates.

Table 1 Tulsa Pre-K Effects on Grade Retention

All	Female	Male
−0.0915***	−0.0617**	−0.1082***
(0.0197)	(0.0275)	(0.0278)
White	Black	Hispanic
−0.0642**	−0.0990***	−0.1632***
(0.0299)	(0.0358)	(0.0485)
Free	Reduced	Paid
−0.0989***	−0.1397**	−0.0481
(0.0255)	(0.0562)	(0.0345)
Full-day	Half-day	
−0.0617**	0.0102	
(0.0251)	(0.0254)	

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Standard errors in parentheses. The reported coefficients are, for the indicated group, for the effects of Tulsa Pre-K on retention in grade as of when the student would have been in ninth grade, for kindergarten students in 2006–2007 who did or did not attend pre-K in 2005–2006. Coefficients come from a linear probability model and hence can be interpreted straightforwardly as the change in the probability of grade retention for a kindergarten student who attended Tulsa pre-K versus a kindergarten student who did not participate in either Tulsa pre-K or Head Start. Regression includes large numbers of controls and propensity score weighting, as described in the text and more fully in Phillips, Gormley, and Anderson (2016). All effects are for students in either full-day or half-day pre-K, except for results in bottom rows.

Table 2 Grade Retention Effects on Earnings by Age, across Subgroups

Age	All	Female	Male	White	Black	Hispanic	Free	Reduced	Paid
18	−671.2***	−741.6***	−629.8*	−1,089***	−222.3	−619.2*	−825.7**	−231.0	−628.2*
	(215.0)	(228.6)	(336.7)	(283.8)	(545.2)	(347.2)	(324.2)	(584.8)	(331.0)
	3,866	*1,894*	*1,972*	*871*	*728*	*2,137*	*991*	*432*	*2,443*
19	−1,405***	−1,474***	−1,262***	−1,801***	−2,110***	−469.6	−1,018**	−2,537***	−1,332***
	(320.2)	(407.3)	(479.3)	(426.1)	(552.1)	(621.4)	(474.0)	(850.9)	(506.5)
	3,581	*1,786*	*1,795*	*825*	*642*	*1,987*	*915*	*404*	*2,262*
20	−1,646***	−1,925***	−1,302**	−2,015***	−2,416***	−718.4	−1,345*	−3,074***	−1,486**
	(455.1)	(594.4)	(663.3)	(609.3)	(930.5)	(847.3)	(694.7)	(1,133)	(711.8)
	3,455	*1,698*	*1,757*	*776*	*620*	*1,948*	*882*	*383*	*2,190*
21	−2,388***	−2,951***	−1,820**	−2,759***	−3,994***	−880.6	−3,921***	−3,687**	−612.4
	(543.6)	(674.1)	(822.2)	(806.2)	(1,093)	(938.1)	(739.8)	(1,493)	(878.4)
	3,417	*1,696*	*1,721*	*753*	*615*	*1,928*	*849*	*376*	*2,192*
22	−3,365***	−4,919***	−1,896**	−2,386**	−4,970***	−2,859***	−4,079***	−3,986***	−2,386**
	(629.1)	(757.7)	(953.7)	(1,031)	(1,261)	(1,006)	(946.9)	(1,520)	(994.2)
	3,412	*1,708*	*1,704*	*745*	*639*	*1,908*	*868*	*372*	*2,172*
23	−3,330***	−5,345***	−1,418	−3,361***	−5,455***	−1,592	−4,339***	−4,452**	−1,894*
	(753.2)	(926.5)	(1,121)	(1,229)	(1,330)	(1,298)	(985.2)	(1,828)	(1,267)
	3,328	*1,640*	*1,688*	*736*	*601*	*1,886*	*851*	*365*	*2,112*
24	−3,825***	−5,206***	−2,524**	−3,881***	−5,433***	−2,310*	−4,758***	−4,479**	−2,610*
	(819.1)	(968.4)	(1,232)	(1,235)	(1,462)	(1,439)	(1,079)	(1,849)	(1,390)
	3,401	*1,689*	*1,712*	*746*	*603*	*1,944*	*856*	*380*	*2,165*
25	−5,283***	−5,745***	−4,851***	−5,494***	−6,800***	−3,207**	−6,489***	−4,204	−4,265***
	(915.1)	(1,237)	(1,293)	(1,359)	(1,783)	(1,564)	(1,117)	(2,972)	(1,507)
	3,511	*1,733*	*1,778*	*774*	*645*	*1,977*	*888*	*372*	*2,251*
26	−6,457***	−6,098***	−6,476***	−5,690***	−6,565***	−5,921***	−6,767***	−7,302***	−5,658***
	(872.2)	(1,121)	(1,263)	(1,366)	(1,608)	(1,477)	(1,217)	(2,009)	(1,432)
	3,571	*1,779*	*1,792*	*805*	*644*	*2,015*	*901*	*407*	*2,263*
27	−5,421***	−5,184***	−5,320***	−4,085***	−4,607**	−5,352***	−4,314***	−4,780**	−6,170***
	(979.4)	(1,232)	(1,434)	(1,429)	(1,979)	(1,637)	(1,288)	(2,332)	(1,633)
	3,655	*1,816*	*1,839*	*837*	*672*	*2,031*	*953*	*398*	*2,304*
28	−5,537***	−4,874***	−5,580***	−3,147*	−3,528*	−6,934***	−3,798**	−8,021***	−6,369***
	(1,121)	(1,413)	(1,651)	(1,775)	(2,099)	(1,853)	(1,507)	(2,830)	(1,833)
	2,962	*1,510*	*1,452*	*655*	*555*	*1,658*	*765*	*324*	*1,873*
29	−7,745***	−7,939***	−7,284***	−3,719*	−10,014***	−8,399***	−6,216***	−8,634***	−8,614***
	(1,217)	(1,332)	(1,868)	(1,933)	(2,153)	(2,065)	(1,623)	(2,825)	(2,006)
	2,907	*1,479*	*1,428*	*680*	*538*	*1,600*	*751*	*315*	*1,841*
30	−8,678***	−7,770***	−9,011***	−9,162***	−8,131***	−7,118***	−6,885***	−11,734***	−9,230***
	(1,364)	(1,581)	(2,127)	(2,060)	(2,226)	(2,553)	(1,722)	(3,192)	(2,348)
	2,196	*1,099*	*1,097*	*493*	*412*	*1,217*	*563*	*241*	*1,392*
31	−8,451***	−7,352***	−9,030***	−2,689	−6,705*	−12,751***	−4,591*	−9,223**	−10,627***
	(1,859)	(1,971)	(3,059)	(3,269)	(3,404)	(3,031)	(2,905)	(3,575)	(2,824)
	1,363	*727*	*636*	*345*	*244*	*738*	*345*	*153*	*865*

NOTE: N in italics. * p < 0.1; ** p < 0.05; *** p < 0.01. Robust standard errors in parentheses. Outcome is earnings from wages/salary/commissions/tips from all jobs before taxes. Group effects are estimated by limiting the sample to that group and are not differential. Estimates control for gender, race/ethnicity, urbanicity, birth cohort, ELL status, free/reduced-price lunch status, single mother/two-parent family, and parental education. Full results are available on request.
SOURCE: Authors' estimates.

Table 3 Grade Retention Effects on Crimes Committed by Type, across Subgroups

Crime type	All	Female	Male	White	Black	Hispanic	Free lunch	Reduced lunch	Paid lunch
Violent crimes	0.0806**	0.0388	0.103*	0.0484	0.0570	0.119***	0.110	0.0519	0.0582*
	(0.0344)	(0.0296)	(0.0570)	(0.0809)	(0.0623)	(0.0436)	(0.0698)	(0.0860)	(0.0340)
Property crimes	0.0721**	0.0201	0.109*	0.0451	0.100	0.0772	0.0923	0.0203	0.0670
	(0.0343)	(0.0252)	(0.0586)	(0.0686)	(0.0798)	(0.0483)	(0.0656)	(0.0764)	(0.0410)
Observations	4,324	2,176	2,148	1,003	810	2,377	1,129	478	2,717

NOTE: * p < 0.1; ** p < 0.05; *** p < 0.01. Robust standard errors in parentheses. Outcome is charges for violent or property crimes between 1998 and 2013, as reported by respondents. Violent crimes are assault (including rape and murder) and robbery. Property crimes are burglary and theft. Subgroup effects are estimated by limiting the sample to the group of interest and are not differential. Estimates control for gender, race/ethnicity, urbanicity, birth cohort, ELL status, free/reduced-price lunch status, single mother/two-parent family, and parental education. Full results are available on request.
SOURCE: Authors' estimates.

Table 4 Tulsa Pre-K Costs, Alternative Estimates

Source of cost figures	Half-day pre-K estimated costs per student ($)	Full-day pre-K estimated costs per student ($)
IWPR	4,508	8,255
NIEER	5,039	9,372
Tulsa	5,297	9,838

NOTE: All figures are in 2013 U.S. dollars. **See text for derivation.**

Table 5 Summary Benefit-Cost Analysis of Tulsa Pre-K, Based on Projecting Grade Retention Results to Earnings Gains and Crime Reductions over the Life-Cycle

Group	Earnings benefit ($)	Crime benefit ($)	Total benefits (earnings gains plus crime reductions) ($)	Pre-K costs ($)	Net benefits ($)	BC ratio	IRR (%)
Overall	14,415	2,963	17,378	8,277	9,102	2.10	5.3
Female	8,168	645	8,813	8,252	561	1.07	3.2
Male	18,420	4,854	23,274	8,299	14,976	2.80	6.3
Black	10,237	3,663	13,900	9,136	4,764	1.52	4.4
Hispanic	19,062	3,763	22,825	8,703	14,123	2.62	6.2
White	9,751	2,883	12,634	7,203	5,431	1.75	4.8
Free lunch	13,072	3,898	16,969	8,776	8,194	1.93	5.2
Reduced lunch	24,577	2,658	27,235	7,942	19,293	3.43	6.9
Paid lunch	7,818	996	8,814	7,025	1,789	1.25	3.6
Full-day	9,721	1,998	11,719	9,838	1,881	1.19	3.5
Half-day	(1,607)	(330)	(1,937)	5,297	(7,235)	−0.37	NA

NOTE: Table reports present value of benefits and costs per Tulsa pre-K participant, overall or in subgroups. Present value is measured in 2013 U.S. national prices. Present value is calculated based on 3 percent real annual discount rate. Derivation of earnings benefits, crime benefits, and costs are discussed in text. Net benefits are simply equal to total benefits measured in this study (earnings and crime reduction benefits) minus pre-K costs. Benefit-cost ratio is simply a ratio of total benefits to costs. IRR is "internal rate of return," which is defined as real discount rate at which benefits just equal costs.

Table 6 Robustness Checks: How Tulsa Pre-K B/C Ratio Varies with Different Estimates, Assumptions, and Methodologies

Group	Baseline	Low end of confidence interval	High end of confidence interval	Zero earnings growth	Low social costs of crime	High social costs of crime	Using test scores to project earnings benefits
Overall	2.10	1.21	2.99	1.59	1.96	2.80	2.25
Female	1.07	0.14	2.00	0.79	1.03	1.19	1.86
Male	2.80	1.39	4.22	2.14	2.58	4.01	2.49
Black	1.52	0.44	2.60	1.22	1.38	2.40	0.01
Hispanic	2.62	1.10	4.15	2.02	2.45	3.58	4.27
White	1.75	0.15	3.36	1.35	1.58	2.40	3.68
Free Lunch	1.93	0.96	2.91	1.51	1.76	2.79	2.09
Reduced Lunch	3.43	0.73	6.13	2.54	3.30	4.02	0.61
Paid Lunch	1.25	−0.51	3.02	0.93	1.20	1.55	3.77
Full-day	1.19	0.24	2.14	0.90	1.11	1.59	1.00
Half-day	−0.37	−2.15	1.42	−0.28	9,−0.34	−0.49	−0.25

NOTE: Baseline benefit-cost ratio from Table 5. Confidence intervals are 95 percent confidence interval based on imprecision in estimating effects of pre-K on retention. Zero earnings growth considers if real earnings do not show any secular real growth. Low and high social costs of crime use various estimates from research literature. Test score projections use effects on eighth grade test scores to predict future earnings benefits. For more details, see text and Appendix C.

Table 7 Comparing This Study's Results with Previous Studies

Study	Program	Methodology	Earnings benefit	Crime benefit	Other benefits	Total benefits	Pre-K costs	BC ratio	BC ratio with only earnings and crime benefits
This study	Tulsa	Projected from middle school grade retention	14,415	2,963	0	17,378	8,277	2.10	2.10
BGA 2012	Tulsa	Projected from kindergarten tests	29,866	0	0	29,866	9,339	3.20	3.20
Belfield et al.	Perry	Observed adult outcomes plus projections	83,834	222,782	11,449	318,065	19,704	16.14	15.56
Heckman et al.	Perry	Observed adult outcomes plus projections	89,794	76,867	9,236	175,897	20,442	8.60	8.15
Reynolds et al.	CPC	Observed adult outcomes plus projections	30,916	45,513	22,415	98,845	9,123	10.83	8.38
Karoly & Bigelow	California universal	Scaled back CPC effects	9,154	2,723	3,883	15,761	5,003	3.15	2.37

NOTE: This study's results come from overall results in Table 5. Other results derived from studies cited in text, and all are overall averages from the cited study. All benefits and costs are expressed in present value terms, as of age 4 for the participant, in 2013 U.S. dollars. Benefits and costs are per pre-K participant. BC ratios are "benefit-cost" ratios, and are ratios of present value of benefits to present value of costs, and can be derived from previous columns in table.