

2011

Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups

Timothy J. Bartik

W.E. Upjohn Institute, bartik@upjohn.org

William Gormley

Georgetown University

Shirley Adelstein

Georgetown University

Upjohn Institute working paper ; 11-176

****Published Version****

[Economics of Education Review](#) 31(6) (2012): 1143-1161

Citation

Bartik, Timothy J., William Gormley, and Shirley Adelstein. 2011. "Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups." Upjohn Institute Working Paper 11-176. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. <https://doi.org/10.17848/wp11-176>

This title is brought to you by the Upjohn Institute. For more information, please contact ir@upjohn.org.

Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups

Upjohn Institute Working Paper No. 11-176

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

William Gormley
Georgetown University

Shirley Adelstein
Georgetown University

July 15, 2011

ABSTRACT

This paper estimates future adult earnings effects associated with a universal pre-K program in Tulsa, Oklahoma. These informed projections help to compensate for the lack of long-term data on universal pre-K programs, while using metrics that relate test scores to valued social benefits. Combining test-score data from the fall of 2006 and recent findings by Chetty et al. (forthcoming) on the relationship between kindergarten test scores and adult earnings, we generate plausible projections of adult earnings effects and a partial cost-benefit analysis of the Tulsa pre-K program. We find substantial projected earnings benefits for program participants who differ by income and by program dosage. The dollar effects and benefit-cost ratios are similar across groups, with benefit-to-cost ratios of approximately 3 or 4 to 1. Because we only consider adult earnings benefits, actual benefit-cost ratios are likely higher, especially for disadvantaged children.

JEL Classification Codes: I210, J240, I240

Key Words: Pre-K, Preschool

Acknowledgments: The authors thank Wei-Jang Huang and Claire Black for their assistance. We also appreciate comments from seminar participants at the Upjohn Institute.

Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups

Timothy J. Bartik
Upjohn Institute

William Gormley
Georgetown University

Shirley Adelstein
Georgetown University

July 15, 2011

ABSTRACT

This paper estimates future adult earnings effects associated with a universal pre-K program in Tulsa, Oklahoma. These informed projections help to compensate for the lack of long-term data on universal pre-K programs, while using metrics that relate test scores to valued social benefits. Combining test-score data from the fall of 2006 and recent findings by Chetty et al. (forthcoming) on the relationship between kindergarten test scores and adult earnings, we generate plausible projections of adult earnings effects and a partial cost-benefit analysis of the Tulsa pre-K program. We find substantial projected earnings benefits for program participants who differ by income and by program dosage. The dollar effects and benefit-cost ratios are similar across groups, with benefit-to-cost ratios of approximately 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 the short-term and long-term effects of such programs, for program beneficiaries and for society as a whole. We now have substantial evidence on the short-term effects of targeted (Reynolds, Temple, Ou, et al. 2011; Schweinhart et al. 2005) and universal (Gormley, Phillips, and Gayer 2008; Henry et al. 2003) programs and the long-term effects of targeted programs (Heckman, Moon, et al. 2010; Reynolds, Temple, Ou, et al. 2011; Reynolds, Temple, White, et al. 2011),, but limited evidence on the long-term effects of universal programs (Karoly and Bigelow 2005). Because the first universal pre-K program in the United States was not established until 1997, we will need to wait a decade or more before we can estimate the consequences of universal pre-K for adults in their 20s and beyond. Even when such long-term data are available on universal pre-K programs, analysts will face the challenge that universal pre-K programs lack experimental data, unlike at least some targeted pre-K programs.

It is possible, however, to make some informed projections, using recent data from Oklahoma's universal pre-K program, which serves both disadvantaged and middle-class children. Using information from recent work by Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (forthcoming) on the link between early test scores and adult earnings, we estimate future earnings for children who did and did not enroll in pre-K, including children who were and were not eligible for a free school lunch. These projections, though not without their limitations, enable us to measure some,

though not all, of the long-term benefits of a high-quality pre-K experience to children from different socioeconomic strata.

These projected earnings effects also allow us to more reliably compare the effects of high-quality pre-K for children from different socioeconomic groups. Without further assumptions, test-score metrics are arbitrary. Therefore, the existing research literature on the short-term test-score effects of high-quality pre-K only allows for metric-dependent and hence arbitrary comparisons of benefits for disadvantaged versus middle-class children. By making additional, informed assumptions about links between test scores and future earnings, we are able to project earnings effects for children from different income groups, and thereby compare effects across income groups based on an important component of social benefits.

Using data collected on children who were just 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. The results of this analysis suggest 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. Furthermore, in each case the benefit-cost ratio well exceeds 1, even though we only consider one particular benefit—adult earnings benefits—in this analysis. When earnings effects are considered relative to expected baseline earnings for each of the income groups, we also find that the percentage effects on future adult earnings are largest for lower-income children. These informed projections allow us to move beyond the

metric of test-score gains to provide some idea of the long-term social benefits of the Tulsa pre-K program.

The remainder of this paper is organized as follows. In section 2, we discuss the limitations of test scores as a metric for assessing program benefits and making meaningful 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, and their applicability in the context of the Tulsa pre-K program. In section 3, we describe the Tulsa pre-K program and present program effects using average test score percentile as our outcome. In section 4, we 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, in section 5 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

Should preschool be targeted at disadvantaged groups, or made universally available? The answer to this depends in part on how the benefits of high-quality preschool vary with a child's family income. Many of these benefits, such as higher adult earnings for former preschool participants, only occur in the long term.

However, good long-term studies of preschool have been confined to disadvantaged groups. For example, the best-known high-quality studies, such as studies of the Perry Preschool Program (e.g., Schweinhart et al. 2005), the Abecedarian Project (Campbell and

Ramey 2010), or the Chicago Child-Parent Center Program (Reynolds, Temple, Ou, et al. 2011) focus on disadvantaged children.

Some good short-term studies of preschool do include children from both disadvantaged and more advantaged families (Gormley et al. 2005; Gormley, Phillips, and Gayer 2008; Wong et al. 2008). The most rigorous short-term studies use a methodology called regression discontinuity. This methodology, to be explained further below, allows us to estimate the effects of preschool on test scores at kindergarten entrance. This methodology yields rigorous estimates without a random assignment experiment, such as that used in recent studies of Head Start (Puma et al. 2005, 2010).

Metric Challenges

But effects on kindergarten test scores have some serious limitations as a metric. Without additional information, it is unclear how kindergarten test-score effects are related to program benefits, such as adult earnings.

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

For example, consider the following pattern of test-score effects of preschool: First, suppose that children eligible for a free lunch (the lowest income group) go from a test score of 20 at preschool entrance to a test score of 40 at kindergarten entrance. Second, suppose that children eligible for a reduced-price lunch (the next-lowest income group) go from a test score of 30 at preschool entrance to a test score of 50 at kindergarten entrance. Finally, suppose that children who must pay full price for lunch (the highest income of these three groups) go from a test score of 40 at preschool entrance to a test score of 60 at kindergarten entrance.

Under these assumptions, who has gained the most in achievement from preschool entrance to kindergarten entrance? 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 a gain from 40 to 60 points, has the same social or economic benefits, or has the same meaning in terms of achievement gains. In percentage terms, the achievement gains are greater for the lower-income groups: a 100 percent gain for the free-lunch children, a 67 percent gain for the reduced-price-lunch children, and a 50 percent gain for the full-price-lunch children. But without further information, comparing these percentage gains is meaningless.

To put this another way, any monotonic transformation of the test-score metric conveys the same information. There is some monotonic transformation of test scores that will yield any pattern of absolute and percentage test-score gains across the three income groups. As Kevin 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, schools, or programs.

Comparing the value of test-score gains requires a judgment as to how test-score gains are related to something we value. Ideally, we would like to relate test-score gains to some comprehensive measure of social benefits. In the real world, we would perhaps settle for relating test-score measures to at least one important component of social benefits.

A Possible Solution

Fortuitously, recent research by Chetty et al. (forthcoming) suggests the shape of the relationship between early test scores and adult earnings. Adult earnings gains are a major component of the social benefits of many educational programs, including high-quality preschool. For example, 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 some welfare benefits.)¹ 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 In contrast to Heckman, Moon, et al.’s results, Schweinhart et al. (2005) find that earnings benefits are only 23 percent of total social benefits for Perry Preschool. But this is because Schweinhart et al. attach much higher valuations per averted crime to the effects of Perry Preschool in reducing crime. The estimated earnings benefits in Schweinhart et al. are only 5 percent lower per participant than in Heckman, Moon, et al.’s results (See Heckman, Moon, et al. [2010], Table 8). Furthermore, because the crime benefits accrue to the rest of society, not program participants, Schweinhart et al.’s results imply that more than 100 percent of the program’s benefits for former participants are due to increased earnings. Rolnick and Grunewald (2003) also find that the earnings benefits of Perry Preschool are only 28 percent of total social benefits (see Heckman, Moon, et al., Table 8). This is in part due to the higher valuation they assign to the program’s effects in reducing crime, but also due to their estimate of earnings effects of the program being 44 percent lower per participant than is true in Heckman, Moon, et al. However, over 100 percent of the Perry Program’s participant benefits in Rolnick and Grunewald’s calculations are due to increased adult earnings.

In both of these studies, the benefits of high-quality preschool are mainly adult earnings gains and lower crime, of which the former is the main benefit to participants.

Chetty et al.'s research is based on 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 variations in 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 IRS administrative data. Chetty et al. find an intriguingly simple relationship between one particular metric for test scores 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. For each test, 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. Chetty et al. mention that this particular test-score metric has a precedent in Krueger's (1999) scaling of test scores in his analysis of the STAR study.

Chetty et al. relate this percentile metric to adult earnings—specifically, average annual earnings from age 25 to 27. Chetty et al. find that this percentile metric is linearly related to adult earnings (Chetty et al., forthcoming, Fig. 1). The relationship seems to be highly linear even in the tails of the distribution of test scores. In other words, an increase in test scores of 10 percentiles gives approximately the same dollar increase in adult earnings, regardless of the initial percentile rank.

One of the central questions investigated by Chetty et al. is 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., forthcoming, Appendix Table 13, col. 1).² This measure controls for the individual student's demographic characteristics, and it attempts to isolate the impact of an induced change in learning conditions. We use this approach to estimate the effects of Tulsa's pre-K program on adult earnings.

Potential Drawbacks

Extrapolating Chetty et al.'s impact estimates to studies of preschool raises some legitimate questions. First, one might wonder whether any initial effects of pre-K on test scores will fade. If fade-out occurs, and persists into lower effects on adult earnings, then extrapolating Chetty et al.'s results to preschool is building a house on sand. We know from the recent Head Start study that Head Start's effects on test scores fade considerably over time (Puma et al. 2010). Even for other pre-K programs, with larger initial effects on test scores, a meta-analysis by Camilli et al. (2010) shows that test-score effects (measured in effect-size units) of pre-K tend to fade, compared to effects at age 5, by 62 percent by age 10.

² Chetty et al. (forthcoming) 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 for the school district. Conversions are done using the national Consumer Price Index for all urban consumers from the U.S. Bureau of Labor Statistics.

However, Chetty et al. also show that even when effects of early interventions on test scores fade out during the K–12 years, these interventions may still have large effects on adult earnings. Specifically, they estimate that improvements in kindergarten class quality, after initial strong effects on test scores at the end of kindergarten, by fourth grade no longer have effects on test scores that are statistically or substantively significant (Chetty et al., Fig. 6). By fourth grade, the test-score effects of kindergarten class quality fade enough so that the predicted effects on adult earnings are less than 10 percent of what would be predicted based on kindergarten test scores. But despite this test score fade-out, by the time individuals are in their mid-20s, the actual, observed effects of kindergarten class quality are similar to what was predicted based on kindergarten test-score effects. The K–12 test score fade-out is a misleading indicator of adult earnings effects.

Similar fade-out of test score effects, and reemergence of effects in adulthood, have been found in some studies of pre-K programs. For example, Deming (2009) finds that Head Start's test-score effects at ages 5 and 6 faded by more than 60 percent (in effect-size units) by ages 11–14 and were no longer statistically significant. However, Head Start's effects on adult outcomes (at age 19 and above) were again statistically significant, and were of similar magnitude (in effect-size units) to the initial test-score effects at ages 5 and 6.

What is going on here? One hypothesis is that the initial effects of early interventions on so-called hard skills, such as literacy and math test scores, proxy for initial effects on both hard skills and so-called soft skills. These soft skills include social skills such as ability to get along with peers and authority figures such as teachers, as well as character skills such as self-

confidence and ability to plan and defer gratification. As James Heckman and his colleagues have hypothesized, this early development of both hard skills and soft skills may lead to further growth in K–12 of hard skills and soft skills (Heckman, Malofeeva, et al. 2010; Heckman, Stixrud, and Urzua 2006). However, the persistence of the effects of soft skills development may be greater than is manifested in later school test scores. Ultimately, these greater long-run soft skills, along with some effects on hard skills, may explain the long-run effects on adult earnings. In sum, early success in hard skills may be associated with greater self-confidence and social skills, which are reflected in higher adult earnings, even when a considerable portion of the initial effects on hard skills fades.

This discussion raises a second question about extrapolating the Chetty et al. results to Tulsa’s pre-K program. 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 persistence of test-score effects. Perhaps the two policies will have different accompanying changes in students that will lead to different effects on adult earnings. For example, perhaps the accompanying change in soft skills will differ across the two policies. As an extreme example, suppose the test-score impacts are generated by very narrow teaching to the test. In that case, one could imagine that this policy might have smaller effects on adult earnings than other policies that yield similar effects on test scores. As we will discuss in section 3, however, the available evidence indicates that Tulsa’s pre-K program is not a narrow program focused only on test-score improvements. The program has broad benefits, including benefits in improved social skills.

A third issue is that preschool impacts on test scores are not necessarily estimated at the end of kindergarten, which is the earliest year for which Chetty et al. link test scores to adult earnings. For example, in this current study, we are measuring preschool impacts on test scores as of the start of kindergarten. However, a defense for using other years is that Chetty et al. also find that test scores measured in percentile ranks are highly correlated for different grade levels. In addition, 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., Appendix Tables 4 and 5). Test-score percentile has somewhat higher dollar effects on adult earnings for test scores measured in percentile terms at the end of sixth through eighth grades. The similarity across early years in the relationship between test-score percentiles and adult earnings makes it plausible that the end of kindergarten/adult earnings relationship will be similar to the start of kindergarten/adult earnings relationship or the end of preschool/adult earnings relationship.

A fourth issue with extrapolating Chetty et al.'s results to predict earnings effects of preschool is differences in local labor markets. One can imagine that the link between early test scores and adult earnings in Tennessee, the source of Chetty et al.'s data, might differ from the link in Tulsa or in other local labor markets. (Chetty et al. include adult earnings wherever earned, but as the sample went to school in Tennessee, they are obviously more likely to stay as adults in Tennessee than is true of individuals who went to school in other states.) The types of jobs, and their wage rates and skill requirements, might differ greatly across local labor markets. Although this is a legitimate concern, migration of even a minority of workers and

employers across local labor markets should result in changes in market wages that limit the differentials across local labor markets in the relationship between adult earnings and adult skills (Marston 1985). If the cross-labor-market differences in the relationship between adult skills and earnings are limited by migration, and 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.

In sum, there are some 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 and certain. However, there are also good reasons to think that for a pre-K program that has demonstrated effects on both hard skills and soft skills, early effects on test scores might provide rough estimates of future earnings effects that can suggest plausible magnitudes of social benefits. The question we now turn to briefly is whether early test-score gains actually do predict future earnings effects for two pre-K programs for which we have actual measures of both early test-score effects and later adult earnings effects.

Evidence from Ypsilanti and Chicago

To see whether Chetty et al.'s approach might provide reasonable estimates for adult earnings gains from pre-K, we will now consider evidence from the Perry Preschool Program and the Chicago Child-Parent Center Program. These two programs were chosen because they are the only two high-quality preschool programs with good evidence on adult earnings in the

mid-20s or later.³ These two preschool programs also have good evidence on these programs' near-term effects on test scores.

For each program, we consider reported data for how the program affected test scores soon after program completion. These test-score effects are translated into an average program effect on test scores measured in percentiles. Based on Chetty et al.'s data on dollar earnings effects per percentile, and the mean annual earnings at ages 25 to 27 in Chetty et al.'s sample, we calculate a percentage effect on adult earnings. Percentage effects on earnings are used rather than dollar effects because the preschool programs' estimated effects on adult earnings are at a wide variety of ages, not ages 25 to 27. It seems reasonable that over the life cycle, earnings effects will tend to go up or down with baseline earnings.⁴ In addition, the earnings metrics used in the various studies differ, which suggests that percentage comparisons are more appropriate.⁵ These test-score-predicted percentage effects on adult earnings are compared with more direct estimates of these programs' effects on adult earnings, based on adult outcomes.

As Table 2.1 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

³ The Abecedarian program as of now has research that uses data on adult outcomes through age 21. However, at this age, many former child participants are still involved in postsecondary education. This is particularly true for participants most favorably affected by the Abecedarian program. Therefore, the available Abecedarian data do not provide as complete a view of effects on adult earnings as are provided by the CPC program and the Perry program.

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 gains predicted by test score percentile effects in preschool or kindergarten. However, assuming constant percentage effects is better than assuming constant dollar effects.

⁵ Chetty et al. (forthcoming) 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, which we would expect to lead to larger dollar effects. Reynolds, Temple, Ou, et al. (2011) use a measure of median earnings within particular income groups, which seems likely to lead to lower dollar effects, as the median value for the lowest income group is zero and the median value for the highest income group is \$30,000 (Reynolds, Temple, Ou et al. 2011, appendix).

findings suggest that short-term test-score effects can be used to provide ballpark estimates of adult earnings effects.⁶

Of course, it would be preferable to actually have random assignment evidence on test-score effects from preschool for a wide variety of income groups, together with random assignment evidence on adult earnings effects for a wide variety of income groups. But acquiring such information would be expensive. The issue is not just monetary costs, but also social opportunity costs. Waiting around for another 25 years for experimental evidence on the adult earnings effects, and other adult outcome effects, of universal preschool for different income groups has a large opportunity cost. This opportunity cost is the possible foregone benefits from not implementing such programs now, compared to waiting 25 or more years for more data. These costs justify making great efforts to see what we can estimate from the available data on different income groups' short-term test-score effects.

Finally, we note that, in addition to Chetty et al.'s paper, there is a precedent for our use of test scores to predict future earnings effects and social benefits of early interventions. In Krueger's (2003) early analysis of the cost and benefits of Tennessee's class size reduction experiment, he heavily relied on estimates of how early elementary test-score improvements would be reflected in adult earnings effects. Krueger's benefit-cost analysis used estimates from Currie and Thomas (1999) of how reading and mathematics test scores at age 7 are

⁶ Early test scores and later earnings results also give similar results if we assume, less realistically, that preschool interventions or test-score improvements give the same dollar increase in earnings at all ages, and regardless of exactly how earnings are measured in the particular study. The Perry test-score percentile effects at the end of preschool predict annual adult earnings gains of \$2,359, measured in fiscal year 2005–06 dollars. Heckman, Moon, et al.'s (2010) numbers imply average annual total compensation gains of \$3,322, in 2005–06 dollars, averaged across ages 19 to 40. This figure is based on direct observations from the Perry data, not extrapolation. The CPC test-score percentile effects in kindergarten predict average annual adult earnings gains of \$1,155, measured in fiscal year 2005–06 dollars. Reynolds, Temple, Ou, et al.'s (2011) measure of effects on median earnings within income groups implies effects of \$754 in year 2005–06 dollars.

related to adult earnings at age 33, using British data. The Currie and Thomas estimates of the effects of early test scores on later earnings are similar in magnitude to the Chetty et al. estimates.⁷

3. Estimating the Percentile Test-Score Effects of Tulsa's Pre-K Program

Suppose we want to use Chetty et al.'s results to predict the adult earnings effects of a pre-K program on different income groups. To do so, what we need is a pre-K program that is truly broad in scope, with data allowing for a reliable estimate of kindergarten test-score effects for different income groups, and with the ability to measure test-score effects with a percentile metric. Fortunately, we have all of these elements in the Tulsa Public Schools pre-K program. This section of the paper uses this program to estimate percentile effects on entering kindergarten test scores.

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 who wish to enroll. Most school districts have chosen to participate, including Tulsa, the largest school district in the state. Although enrollment is voluntary, the overwhelming majority of parents have chosen to enroll their four-year-olds. Oklahoma now leads the nation in preschool access, with over 70 percent of four-year-olds enrolled (Barnett et al. 2010).

⁷ A discussion and comparison of the Currie and Thomas versus the Chetty et al. estimates is provided in Bartik (2011), endnote 3, starting on p. 349.

The Tulsa pre-K program is a high-quality program. Every teacher has a B.A. degree, is early-childhood-certified, and earns the same salary as other public school teachers, as required by state law. A 10-to-1 child/staff ratio is maintained. A systematic 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 (as measured by productivity, instructional learning formats, concept development, and quality of feedback) is also higher in Tulsa than elsewhere (Phillips, Gormley, and Lowenstein 2009).

The Tulsa pre-K program is clearly not simply focusing narrowly on improving test scores on reading and math tests. For example, previous analyses of the Tulsa pre-K program have found some statistically significant but modest positive effects on social-emotional development (Gormley et al., forthcoming). Suppose, as some researchers have hypothesized, that effects of preschool on soft skills are an important mechanism for these programs' long-run effects (Heckman, Malofeeva, et al. 2010). Then the broad estimated effects of the Tulsa pre-K program reduce the possibility that measured effects on reading and math skills will understate later effects on earnings.

In short, the Tulsa pre-K program is a better-than-average pre-K program that reaches large numbers of students. In this sense, it differs from the justly celebrated Perry Preschool and Abecedarian Project programs, which reached very small numbers of students. It also differs from these programs in its relatively low cost. Whereas the Tulsa pre-K program cost an average of \$4,403 in its half-day version, and \$8,806 for its full-day version (see section 4), the

Perry Preschool Program cost \$17,526 per child, and the Abecedarian Project cost \$39,672 per child in fiscal year 2005–06 dollars (Masse and Barnett 2002, Table 8.2, p. 45; Schweinhart et al. 2005, Table 7.8, p. 148).⁸ On the other hand, the Tulsa pre-K program is similar in costs to the Chicago CPC program, which cost \$5,372 per year for a half-day program that included not only school-year services, but also some summer services.⁹ The Chicago CPC is also a larger-scale program than the Perry or Abecedarian programs, though not of the size and scope of the Oklahoma program.

Tulsa Data

The data available to us are derived from student testing conducted in August 2006. Just prior to the commencement of classes, teachers were instructed to administer three subtests of the Woodcock-Johnson Achievement Test to incoming kindergarten students and to incoming pre-K students. The three subtests, appropriate to relatively young children, were 1) Letter-Word Identification (a measure of prereading skills), 2) Spelling (a measure of prewriting skills), and 3) Applied Problems (a measure of 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, conducted at the same time as the testing. Administrative records specified each child's gender,

⁸ The Perry Preschool program was a two-year program for most participants, so the cost for a one-year program would be lower. However, there is insufficient sample size to estimate the effects of the Perry program for one-year participants, so published findings are all based on combining participants for both one year and two years.

⁹ Data are taken from Reynolds, Temple, White, et al. (2011) but are 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 available benefit-cost analyses indicate that the one-year program had a higher benefit-cost ratio than the two-year program (Reynolds, Temple, White, et al. 2011, Table 5).

date of birth, race/ethnicity, and school lunch status, a surrogate for income. The parent survey, which we 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 and Rubin 2002; Rubin 1987).¹⁰ This method involves creating multiple complete data sets containing 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 common missing data techniques (e.g., Croy and Novins 2005; Rubin 1996; Sinharay, Stern, and Russell 2001).

Estimating Technique

As with previous studies of the Tulsa pre-K program (Gormley and Gayer 2005; Gormley et al. 2005; Gormley, Phillips, and Gayer 2008), we used a regression-discontinuity design to guard against selection bias. We compared incoming kindergarten students who had participated in pre-K the previous year (the treatment group) with incoming pre-K students (the control group). This research design addresses the crucial concern that certain families are

¹⁰ We implemented multiple imputation with five imputes using the Stata ICE program (Royston, Carlin, and White 2009), which generates plausible values using imputation by chained equations (van Buuren, Boshuizen, and Knook 1999). For our analytic sample, the prevalence of 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.

more likely to select into the pre-K program, and these families may have unobservable characteristics that are also related to test scores.

Obviously, the students differ, by one year, in their average age, which is positively related to test outcomes. However, we controlled for the students' date of birth and other covariates. The big advantage of our approach, as opposed to other nonexperimental alternatives, is that both sets of students participated in the program, albeit a year apart on average. This means that both sets of parents affirmatively chose to enroll their students in the program, which should, in principle, eliminate differences in unobservable characteristics associated with program participation, such as parent or student motivation.

Whether a child is in the treatment or comparison group is dependent on his or her date of birth. The state of Oklahoma enforces a strict birthday cutoff for program eligibility in a given year. Thus, 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 are able to compare pre-K entrants who just missed the birthday cutoff with pre-K alumni who just made it.

If the regression discontinuity research design works, then students in the treatment and control groups should have similar demographic characteristics, particularly for observations closest to the cutoff. As Table 3.1 indicates, this was the case. The results in Table 3.1 show the differences in test scores and observable characteristics around the birthday cutoff. These parametric fits reflect predicted differences at the cutoff and are equivalent to testing for a “jump” at the regression discontinuity. 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.

With a regression discontinuity design, one can use different “windows,” or ranges, around the cutoff—e.g., a 12-month window, a 6-month window, a 3-month window, or even a 1-month window. We have chosen to use a 12-month window, to maximize our sample size. Though restricting the analysis to observations closest to the cutoff reduces any potential bias, this also greatly reduces the precision of the estimates. In previous studies, the choice of a window yielded somewhat different effect sizes but did not yield fundamentally different results.

Average Test Percentile Dependent Variable

In contrast to previous studies using the Tulsa data, we have created a different dependent variable, to mirror as closely as possible the approach taken by Chetty et al. (forthcoming). Previous Tulsa studies have utilized raw test scores and have reported results for the three Woodcock-Johnson subtests separately. In this study, we utilize percentile ranks, and we combine the three subtests into one measure. Specifically, we scaled scores for each of the subtests into percentile ranks based on the distribution of test scores in 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. In section 4, we will combine these percentile estimates with Chetty et al.’s results to predict future earnings effects and compare these earnings effects to program costs.

Preliminary Results for Comparison with Previous Studies

Returning to Table 3.1, we observe (as expected) a substantial and highly significant ($p < 0.001$) difference in average test percentile at the cutoff. These results suggest that program participation is associated with a 14.9-percentile-point increase in test scores. Another way to estimate this effect is to use a single equation model in which test scores are regressed on precise date of birth (number of days born before or after the birthday cutoff), an indicator for whether the person was born before the birthday cutoff, and an interaction term between these two variables to allow for different slopes on either side of the cutoff. Table 3.2 presents the results of such a model, including a number of covariates, for the full sample in the 12-month window. Columns 2–4 present similar results for three lunch status subgroups: free lunch, reduced-price lunch, and full-price lunch.

Similar to the results in Table 3.2, for the overall sample, program participation increases test scores by 15.5 percentile points on average, an effect size of 0.89. This effect size, using percentile ranks and combining all three tests, appears somewhat higher than the average effect size for the three tests separately, using raw test scores: 0.98 for Letter-Word Identification, 0.74 for Spelling, and 0.36 for Applied Problems (Gormley et al. 2010), with an overall average effect size across these three tests of 0.69. The reason for the slightly higher effect sizes with this different dependent variable is unclear. Perhaps the averaging involved in creating the percentile variable lowers somewhat the natural population variation or standard deviation, which will tend to raise effect sizes.

As Table 3.2 indicates, effect sizes for pre-K participation are impressive generally, but especially for free-lunch and reduced-price-lunch students. 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 0.57 to 0.98 to 1.21. These higher effect sizes occur in part because the estimated effect on the dependent variable is somewhat higher for free-lunch and reduced-price-lunch students. But the higher effect sizes for low-income groups also occur because the standard deviation of the dependent variable is lower among pre-K entrants in low-income groups—that is, the “natural variation” in the dependent variable is somewhat lower in the lower-income groups. The effect-size methodology can be interpreted as saying that effects that are larger relative to natural variation are in some sense “more important” or at least more difficult to obtain.

These “effect size” variations across income groups are qualitatively similar although quantitatively somewhat larger than previous results for the Tulsa pre-K program (Gormley et al. 2005). These previous results are from the fall of 2003 and are for three different tests. In this previous study, for example, average effect sizes across the three tests were 0.49 for the full-price-lunch group, compared to 0.64 for the free-lunch group. In the current study, effect sizes go up from 0.57 for the full-price-lunch group to 1.21 for the free-lunch group.

Another metric for comparing different income groups is to examine percentage effects on test scores. Compared to where the children start out at entrance to pre-K, by what percentage is pre-K estimated to improve their test-score performance? In other words, this approach divides the estimated test-score effects by the control group means for that particular

income group. By this metric, the percentage test-score effects vary even more by income group. As shown in Table 3.2, percentage effects are 44.1 percent for the full-price-lunch group, increasing to 82.5 percent for the reduced-price-lunch group, and then up to 112.8 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. If what we view as important is percentage effects on a child's test-score gains, then clearly pre-K programs in Tulsa have "more important" effects on low-income children.

These percentage variations across different income groups are qualitatively similar to previous years' results for Tulsa, but even more pronounced quantitatively. For example, consider some percentage results from the fall of 2003 in Tulsa (Gormley 2010). Across the three tests considered, average percentage effects increased from 22 percent for the full-price-lunch group to 38 percent for the free-lunch group. The jump in percentage effects in the current study is far greater, from 44 percent for the full-price-lunch group to 113 percent for the free lunch group.

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

Percentile Test Score Effects Disaggregated by Lunch Status and Half-Day vs. Full-Day Program

We focus more attention in this subsection on the percentile effects. We will use percentile effects in section 4 to calculate adult earnings effects. These adult earnings effects

will then be compared with program costs to obtain a partial estimate of program benefits versus costs.

To allow for a proper comparison with program costs, we need to also 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 there are differences across income groups in relative enrollment in full-day versus half-day pre-K. Specifically, in our sample, 78 percent of the free-lunch pre-K entrants enrolled in full-day pre-K, and 22 percent in half-day pre-K. In contrast, only 45 percent of the full-price-lunch pre-K entrants enrolled in full-day pre-K, versus 55 percent in half-day pre-K. (The reduced-price-lunch pre-K entrant group was in-between, with 66 percent enrolling in full-day pre-K, 34 percent in half-day pre-K.)

These different enrollment patterns are in part 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, one could argue that low-income families will be more in need of full-day programs, both to help their children and to provide free child care, which these families will have more need of, both because of their lower disposable income and because of the likely greater proportion of single mothers among low-income families. Therefore, Tulsa, like many other school districts in

Oklahoma, 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 or needs. Parents are not allowed to select a full-day versus a half-day program at a particular neighborhood school, as each school only offers a particular type of program. However, parents choose whether to participate in the program at all, because although Oklahoma's pre-K program is "universal," it is also voluntary. In addition, there were two Tulsa public schools in 2005–06 that offered some full-day slots to nonneighborhood children, with free transportation. It is certainly plausible that low-income parents may find full-day pre-K programs more appealing, relative to half-day programs, while middle-class parents may be more attracted to half-day programs.¹¹

Regardless of what causes these enrollment differences across income groups, we need to control for these differences to properly compare the benefits and costs of pre-K across income groups. We don't want to conclude that pre-K is more effective for low-income groups simply because a higher percentage of the low-income group enrolls in full-day pre-K programs.

Table 3.3 presents the percentile test-score effect of Tulsa pre-K, disaggregated both by full- versus half-day status and 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 3.2. As the first row of numbers in Table 3.3 makes clear, for a given type of program (full-day versus half-day),

¹¹ On the whole, conversations with Tulsa school officials suggest that most parents prefer full-day pre-K programs. In response to this demand, as of the 2010–11 school year, virtually every Tulsa school now offers only a full-day program. However, this was not the case in the 2005–06 school year, from which our data come, in which about one-third of the students (more precisely, in our data, 31.7 percent) are in half-day programs.

there are no large differences in percentile test effects across the three different income groups. The differences across different income groups are not only statistically insignificant but also substantively insignificant.

However, this is not the case when we look at effect size or percentage effects. Relative to the natural variation in test scores in the control group of pre-K entrants, effect sizes of Tulsa pre-K tend to be considerably greater for the free-lunch group than for the full-price-lunch group. Relative to the average starting scores in the control group of pre-K entrants, the advantage for the free-lunch group over the full-price-lunch group is even greater. Therefore, if we 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 or more important or more difficult to achieve.

However, based on Chetty et al.'s results, the effects on test-score percentiles may be more directly related to one of the main social benefits of pre-K programs, the effects on adult earnings. Section 4 will consider the implications of the Table 3.3 results for the adult earnings effects, as well as the benefits versus the costs of the Tulsa pre-K program for different income groups.

4. Predicting the Future Earnings Benefits of the Tulsa Pre-K Program

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 effects of Tulsa pre-K on adult earnings. These estimated earnings effects are then

compared with predicted future earnings effects of the various income groups and with program costs.

Predicted Future Earnings Benefits

Even without much number-crunching, it can easily be seen that the Chetty et al. percentile effects on adult earnings, and our results on Tulsa percentile effects, imply large effects on the present value of adult earnings. This can immediately be seen because the Chetty et al. results imply an annual earnings effect at ages 25 to 27 of around \$73 per a 1-percentile boost. Given that this would probably go up as this cohort ages and 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 at reasonable interest rates. Given that the Tulsa pre-K program boosted the percentile test scores by 8 to 20 percent, the implied effects are quite large.

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

We then calculated the effect of a 1-percentile early test-score boost at each age. We assumed 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 mean earnings effects for ages 25–27, the years considered by Chetty et al., would exactly match Chetty et al.'s estimate of a \$73.01 boost for each increase in test scores by 1 percentile.

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, compared to percentage effects for former participants in their mid-20s. Heckman et al.'s results imply that Perry Preschool's effects on adult earnings average 8 percent from ages 19 to 27, 26 percent from ages 28 to 40, and 21 percent from ages 41 to 65. The average annual dollar earnings effects of Perry Preschool varies even more with age: In 2005–06 dollars, these average annual dollar earnings effects are \$1,439 at ages 19–27, \$5,621 at ages 28–40, and \$3,596 at ages 41–65. Therefore, assuming a constant percentage effect at later ages than the mid-20s is better than assuming a constant dollar effect. But the true percentage effect does probably grow some over time compared to the mid-20s.

We then discounted these earnings gains to age 4. This discounting is done so that these earnings benefits can later be compared with Tulsa pre-K program costs, which would be incurred when participants are age 4. We used a real discount rate of 3 percent, which is a fairly typical social discount rate. However, we note that since 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. So, for example, the Social Security Trustees assume long-run average real earnings growth of 1.2 percent per year (Board of Trustees 2011, Table II.C1, p. 7). If this occurs, then the real earnings numbers here should be increased by the real earnings increase between 2005–06 and the future year in which these earnings increases occur. An equal-sized increase in the discount rate would offset this secular increase in real earnings.

We end up with a discounted present value of an increase of \$1,502 in adult earnings for a 1-percentile increase in test scores. Multiplying this \$1,502 figure by the estimates in Table 3.3, 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 4.1.

As expected, these estimated earnings benefits of Tulsa pre-K are large. These earnings benefits would justify quite high program costs to achieve these benefits, or many times 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 in absolute terms do not differ much across income groups.

Predicting Earnings Effects for Different Income Groups in Percentage Terms

Although from an economic efficiency perspective the earnings benefits for the different income groups are similar, society may put a greater weight on the earnings gains to lower-

income groups.¹² To get some sense about how these earnings gains compare with the income status of the different income groups, we estimate the baseline future earnings for children in each of the three income groups.

Because the estimated dollar effects are similar across different income groups, a pure efficiency-based benefit-cost analysis would be similar across the three income groups. In other words, for a benefit-cost analysis that simply counts dollar gains and losses, regardless of who receives these gains and losses, Tulsa pre-K would be judged as having similar social benefits for all three income groups. However, the three income groups of children differ in their likely baseline earnings prospects as adults. For example, we would expect the children from families eligible for a free lunch to have lesser baseline earnings prospects as adults than the children from families eligible for a full-price lunch. Even with similar dollar effects across income groups, the different baseline earnings levels imply different percentage earnings impacts. The percentage effects on future adult earnings will be greater for free- and reduced-price-lunch children compared to full-price-lunch children.

Because the baseline earnings levels are lower for children from free lunch–eligible families, a dollar gain from pre-K for this group would be judged by many policy analysts as being worth more socially than a similar dollar gain to someone whose baseline earnings

¹² This can be done directly, although somewhat mechanically, by applying distributional weights that place a greater value on a dollar gain to someone of low income than to someone of high income. There is an extensive literature in benefit-cost analysis discussing the pros and cons of distributional weights (e.g., Johansson-Stenman 2005).

prospects are higher.¹³ The extent to which this is true depends upon how much baseline earnings levels vary, which will affect percentage earnings effects.

Percentage earnings effects may not be the exactly right social weighting of benefits. Percentage earnings effects would only be the exactly right criteria for comparing social benefits if the social weight on a gain to income for a particular income group is proportional to one over the group's baseline expected earnings level. However, even for other social weightings of the earnings gains of different income groups, these percentage earnings gains calculations give some idea of the distributional effects of the Tulsa pre-K program. Furthermore, with baseline earnings levels estimated, policymakers can apply their own social weights to the dollar earnings gains of each income group. Even if policymakers don't want to explicitly use distributional weights, the pattern of percentage earnings gains across income groups give some guidance as to how qualitatively important the pre-K program's effects are to different income groups.

We obviously do not know for certain the baseline future earnings levels of these Tulsa children who participated in pre-K in 2005–06. Among other things, these future earnings have not had a chance to occur yet. Therefore, we will have to make some assumptions to predict those likely future earnings.

To predict future earnings, we isolated records from the American Community Survey of all children ages 4 to 18 who lived in the city of Tulsa and who attended public schools.

¹³ This can be justified from a utilitarian perspective based on diminishing marginal utility of income (Drèze and Stern 1987; Little and Mirrlees 1974; Marshall 1885). Alternatively, we can argue that society ought to be maximizing the prospects for the least-well-off (Rawls 1971). Finally, we can argue that society ought to be emphasizing enhancing basic capabilities for all persons (Nussbaum 2011; Sen 1992), and that this is more likely to be done with greater assistance to low-income groups.

(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 a full-price lunch.

For each of these three groups of children, we looked at the earnings of parents. We calculated for each group of parents the mean earnings, broken down by gender and age. These mean earnings calculations used the child weights in the ACS. Because of the large sample size of the ACS, we ended up with a somewhat surprisingly large sample size: 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.

These 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 of that combination 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 took a simple average across mothers and fathers for each income group to get an average percentage that that income group represents of overall metro-area earnings.

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

However, these are typical earnings for the parents. We would expect some regression to the mean in the child's future earnings as an adult. Specifically, we would expect the child's earnings to on average be between his parents' earnings and overall Tulsa metro-area earnings.

The research literature suggests that a plausible value for the coefficient on $\ln(\text{parent earnings})$, in a regression explaining the natural log of the child's earnings as an adult, might be 0.4 (Chadwick and Solon 2002; Solon 2002). That is, we would expect the natural log of the child's average earnings to 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 the parents' earnings.

If we do 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 ($65.0 = \exp[(0.4) \cdot \ln(34.1) + (1-0.4) \cdot \ln(100.0)]$). 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 discounted present value as of age 4 of Tulsa metro-area average earnings from ages 22 to 66, discounted at 3 percent, and not allowing for any real earnings growth over time (or, alternatively, assuming a discount rate equal to 3 percent plus the average annual growth rate of real earnings), is \$401,833. Using the above percentages, the average earnings of the children from the three income groups can be calculated as adults: free-lunch children, \$261,308 (= 65.0 percent x \$401,833); reduced-price-lunch children, \$343,038; and full-price-lunch children, \$453,430. These baseline earnings figures can then be compared with the estimated dollar effects of Tulsa pre-K to get percentage effects.

These percentage effects are reported in the fifth row of numbers in Table 4.1. These percentage effects do differ considerably across income groups. For example, for a half-day pre-K program, the percentage effects on earnings are almost twice as great for the free-lunch group as for the full-price-lunch group (6.8 percent versus 3.6 percent). For full-day pre-K, the differential in percentage effects between the free-lunch group and the full-price-lunch group is even greater (10.4 percent versus 5.4 percent). These differences are sufficient that it would certainly be reasonable for a policymaker concerned with income inequity, and who therefore valued gains to lower-income groups as more important, to 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 is in large part because of the “regression to the mean” assumptions made about the child’s future adult earnings. The free-lunch-eligible group’s parents have quite low earnings, at

only about one-third of the metro-area average (more precisely, 34.1 percent). Furthermore, these parents' earnings 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). However, these baseline earnings differentials narrow considerably for the child's future earnings. While the parents of the free-lunch children have earnings that are only one-fourth of those of the parents of the full-price-lunch children, 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).

What this illustrates is 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. Obviously a pre-K program cannot know a child's baseline future earnings with any precision, so it is hard to determine perfectly which children will be most in need of this program. While there is considerable economic persistence across generations, there is also considerable upward and downward mobility, as noted by other researchers (Isaacs 2007).

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 gives us an idea of the extent to which the Tulsa pre-K program may be expected to yield earnings benefits for participants in adulthood, and it

suggests that these benefits may be disproportionately larger for participants from more disadvantaged families.

A Partial Benefit-Cost Analysis

These predicted future earnings effects of Tulsa pre-K can also 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 the benefits of Tulsa pre-K. Among the most important benefits that are not included are the social value of whatever reduction in crime might be caused by the program and the cost savings from reduced special-education costs or other remedial education costs in K–12.

To do such 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. This estimate is for all of Tulsa Public Schools; school officials were unable to provide a disaggregation that specified the local share for pre-K.

This local share exceeds the estimated local share for all of Oklahoma for 2005–06 that is provided in the annual reports of the National Institute for Early Education Research (NIEER). NIEER estimates that for each dollar of state aid for pre-K in Oklahoma, local school districts provide 57 cents (Barnett et al. 2006). We chose the larger local-share number for two reasons.

First, we wanted to be conservative in our cost estimates and err on the side of overestimating program costs. Second, we felt that it was plausible that a large urban school district such as Tulsa, with a large property tax base, might provide a larger local share for pre-K 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 or diseconomies of scale from doubling the length of the pre-K day.

Based on these data and these conversations, we concluded that in 2005–06, a half day of Tulsa pre-K cost \$4,403, and a full day of Tulsa pre-K cost twice as much, at \$8,806. 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 4.1.

As can be seen by these benefit-cost ratios, for all income groups and all program-day lengths, the ratio of earnings benefits to costs is much greater than one. That is, even considering only adult earnings benefits, 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, such as possible reductions in crime or special education costs.

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 based on much less precise test-score effects, because of the much smaller sample sizes for the reduced-price-lunch group (see Table 3.3). Therefore, from a pure efficiency perspective, as noted before, and including 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 different income groups, which would change the analysis even from an efficiency perspective.

It is tempting to use the benefit-cost figures contained in Table 4.1 to calculate the incremental benefits versus costs of a child in a given income group moving from a half-day program to a full-day program or from a full-day program to a half-day program. The last row in Table 4.1 presents the incremental benefit-cost ratios of moving a child in each lunch-status (income) group from a half-day program to a full-day program, calculated as the difference in earnings benefits, based on average earnings benefits in each category of program for that income group, divided by the average differences in costs. These results would seem to suggest

that for the free-lunch and full-price-lunch groups, the incremental benefits of switching from a half-day to a full-day program are less than the benefit-to-cost ratio for a half-day program. This is not the case for the reduced-price-lunch group, but, again, these numbers are based on a smaller sample.

It is important to be cautious in interpreting such comparisons. First, these 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, there are reasons to think that this may not be the case. Families participate in full-day versus half-day program types based partly on their own selection into a desired program and partly on decisions by Tulsa Public Schools about whether to place each type of program in a given school. 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. In addition, 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. Any net benefits from additional participation from offering full-day pre-K need to be considered.

5. Conclusion

Because we are estimating future benefits that have not yet occurred, and because we are linking data from two research sites (Tulsa, Okla., and Tennessee), it is important to stress

the limitations of our findings. First, we cannot be sure that kindergarten test scores in Oklahoma will translate into adult earnings in Oklahoma in precisely the same way that kindergarten test scores in Tennessee have translated into adult earnings in Tennessee. One obvious difference is that different job markets could have different returns to different skills. Another difference is that the 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.

Second, we have focused exclusively on earnings, while ignoring possible effects on remedial education and crime. Other studies of the long-term effects of pre-K have found criminal justice effects to be particularly important (Heckman, Moon, et al. 2010; Reynolds, Temple, Ou, et al. 2011; Rolnick and Grunewald 2003; Schweinhart et al. 2005). On the other hand, these other studies have focused exclusively on disadvantaged children. One might expect to see less dramatic effects on crime for the more heterogeneous population served by Oklahoma's universal pre-K program. The same could also be true of remedial education, which, like crime, afflicts more disadvantaged children more grievously. In any event, this omission probably 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.

More plausibly, our focus on short-term “hard skills” can be defended as assuming that short-term effects on hard skills can proxy for the overall short-term effects of the program, as clearly these effects will be linked. 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 and Thomas 1999) that short-term effects on hard skills can be used 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 attentiveness and short-term reductions in timidity have more profound long-term consequences than short-term gains in prereading, prewriting, and premath skills. It is possible that a fuller understanding of the links between all these short-term effects and long-run earnings could lead to somewhat different and 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 that reaches children from both more- and less-advantaged households. It also illuminates benefit and cost differentials for programs that vary in their dosage (half-day versus full-day) and for programs that vary in their beneficiaries (free versus reduced versus no subsidized school lunch). What is most striking about these benefit-cost comparisons is not the modest variations by dosage and child poverty but rather the

striking similarities. For all children, irrespective of income and irrespective of program duration, 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.

In addition to this evidence of substantial returns across income and program-type subgroup, our results also suggest that children from more-disadvantaged families are likely to see the largest relative adult earnings benefits. Although the absolute earnings benefits and benefit-cost ratios are similar across income, 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 are expected to experience percentage gains that are nearly 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 cost-benefit ratios across groups must be considered in light of the fact that, as noted above, this analysis has been limited to adult earnings benefits. In contrast, we have made every effort to account for all of the program costs. 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.

References

- Barnett, W. Steven, Dale J. Esptein, Megan E. Carolan, Jen Fitzgerald, Debra J. Ackerman, and Allison H. Friedman. 2010. *The State of Preschool 2010*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. Steven, Jason T. Hustedt, Laura E. Hawkinson, and Kenneth B. Robin. 2006. *The State of Preschool 2006*. New Brunswick, NJ: National Institute for Early Education Research.
- Bartik, Timothy J. 2011. *Investing in Kids: Early Childhood Programs and Local Economic Development*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds (Board of Trustees). 2011. *The 2011 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. House Document 112-23. Washington, DC: Government Printing Office..
- Camilli, Gregory, Sadako Vargas, Sharon Ryan, and W. Steven Barnett. 2010. "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teachers College Record* 112(3): 579–620.
- Campbell, Frances A., and Craig T. Ramey. 2010. "Carolina Abecedarian Project." In *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration*, Arthur J. Reynolds, Arthur J. Rolnick, Michelle M. Englund, and Judy A. Temple, eds. New York: Cambridge University Press, pp. 76–98.
- Chadwick, Laura, and Gary Solon. 2002. "Intergenerational Income Mobility among Daughters." *American Economic Review* 92(1): 335–344.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. Forthcoming. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4).
- Croy, Calvin D., and Douglas K. Novins. 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, Janet, and Duncan Thomas. 1999. "Early Test Scores, Socioeconomic Status, and Future Outcomes." NBER Working Paper No. 6943. Cambridge, MA: National Bureau of Economic Research.

- Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111–134.
- Drèze, Jean, and Nicholas Stern. 1987. "The Theory of Cost-Benefit Analysis." In *Handbook of Public Economics*, Vol. 2, Alan J. Auerbach and Martin Feldstein, eds. Amsterdam: North-Holland, pp. 909–989.
- Gormley, William T., Jr., 2010. "Small Miracles in Tulsa: The Effects of Universal Pre-K on Cognitive Development." In *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration*, Arthur J. Reynolds, Arthur J. Rolnick, Michelle M. Englund, and Judy A. Temple, eds. New York: Cambridge University Press, pp. 188–198.
- Gormley, William T., Jr., and Ted Gayer. 2005. "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program." *Journal of Human Resources* 40(3): 533–558.
- Gormley, William T., Jr., Ted Gayer, Deborah A. Phillips, and Brittany Dawson. 2005. "The Effects of Universal Pre-K on Cognitive Development." *Developmental Psychology* 41(6): 872–884.
- Gormley, William T., Jr., Deborah A. Phillips, Shirley Adelstein, and Catherine Shaw. 2010. "Head Start's Comparative Advantage: Myth or Reality?" *Policy Studies Journal* 38(3): 397–418.
- Gormley, William T., Jr., Deborah A. Phillips, and Ted Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320(5884): 1723–1724.
- Gormley, William T., Jr., Deborah A. Phillips, Katie Newmark, Kate Perper, and Shirley Adelstein. Forthcoming. "Social-Emotional Effects of Early Childhood Education Programs in Tulsa." *Child Development*.
- Heckman, James J., Lena Malofeeva, Rodrigo Pinto, and Peter A. Savelyev. 2010. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." Unpublished manuscript, University of Chicago Department of Economics.
- Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the High/Scope Perry Preschool Program." *Journal of Public Economics* 94(1-2): 114–128.

- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24(3): 411–482.
- Henry, Gary T., Laura W. Henderson, Bentley D. Ponder, Craig S. Gordon, Andrew J. Mashburn, and Dana K. Rickman. 2003. *Report of the Findings from the Early Childhood Study: 2001–02*. Atlanta: Andrew Young School of Policy Studies, Georgia State University.
- Isaacs, Julia B. 2007. *Economic Mobility of Families across Generations*. Washington, DC: Brookings Institution.
- Johansson-Stenman, Olof. 2005. "Distributional Weights in Cost-Benefit Analysis: Should We Forget about Them?" *Land Economics* 81(3): 337–352.
- Karoly, Lynn A., and James H. Bigelow. 2005. *The Economics of Investing in Universal Preschool Education in California*. Santa Monica, CA: RAND Corporation.
- Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2): 497–532.
- . 2003. "Economic Considerations and Class Size." *Economic Journal* 113(485): F34–F63.
- Lang, Kevin. 2010. "Measurement Matters: Perspectives on Education Policy from an Economist and School Board Member." *Journal of Economic Perspectives* 24(3): 167–182.
- Little, I.M.D., and J.A. Mirrlees. 1974. *Project Appraisal and Planning for Developing Countries*. London: Heinemann.
- Little, Roderick J.A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley and Sons.
- Marshall, Alfred. 1885. *The Present Position of Economics*. London: Macmillan.
- Marston, Stephen T. 1985. "Two Views of the Geographic Distribution of Unemployment." *Quarterly Journal of Economics* 100(1): 57–79.
- Masse, Leonard N., and W. Steven Barnett. 2002. *A Benefit-Cost Analysis of the Abecedarian Early Childhood Intervention*. New Brunswick, NJ: National Institute for Early Education Research.

- Nussbaum, Martha C. 2011. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Harvard University Press.
- Phillips, Deborah A., William T. Gormley, Jr., and Amy E. Lowenstein. 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, Michael, Stephen Bell, Ronna Cook, and Camilla Heid. 2010. *Head Start Impact Study: Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, and Michael Lopez. 2005. *Head Start Impact Study: First Year Findings*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Reynolds, Arthur J. 1995. "One Year of Preschool Intervention or Two: Does It Matter?" *Early Childhood Research Quarterly* 10(1): 1–31.
- Reynolds, Arthur J., Judy A. Temple, Suh-Ruu Ou, Irma A. Arteaga, and Barry A.B. White. 2011. "School-Based Early Childhood Education and Age-28 Well-Being: Effects by Timing, Dosage, and Subgroups." *Science* 333(6040): 360–364.
- Reynolds, Arthur J., Judy A. Temple, Barry A.B. White, Suh-Ruu Ou, and Dylan L. Robertson. 2011. "Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program." *Child Development* 82(1): 379–404.
- Rolnick, Art and Rob Grunewald. 2003. "Early Childhood Development: Economic Development with a High Public Return." Federal Reserve Bank of Minneapolis's *The Region* 17(4): 6–12.
- Royston, Patrick, John B. Carlin, and Ian R. White. 2009. "Multiple Imputation of Missing Values: New Features for MIM." *Stata Journal* 9(2): 252–264.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- . 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91(434): 473–489.

- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: HighScope Educational Research Foundation.
- Sen, Amartya. 1992. *Inequality Reexamined*. New York: Russell Sage Foundation; and Cambridge, MA: Harvard University Press.
- Sinharay, Sandip, Hal S. Stern, and Daniel Russell. 2001. "The Use of Multiple Imputation for the Analysis of Missing Data." *Psychological Methods* 6(4): 317–329.
- Solon, Gary. 2002. "Cross-Country Differences in Intergenerational Earnings Mobility." *Journal of Economic Perspectives* 16(3): 59–66.
- Van Buuren, S., H.C. Boshuizen, and D.L. Knook. 1999. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis." *Statistics in Medicine* 18(6): 681–694.
- Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. 2008. "An Effectiveness-based Evaluation of Five State Pre-kindergarten Programs." *Journal of Policy Analysis and Management* 27(1): 122–154.

Table 2.1 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
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 the following tests: the Stanford-Binet IQ Test, the Leiter International Performance Test (nonverbal intellectual 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. (forthcoming) estimate that a 1-percentile change in test scores increases annual earnings for ages 25–27 by \$73.01. The dollar earnings effect was then divided by Chetty et al.’s estimates of average annual earnings at ages 25–27 to get percentage effects. Perry Preschool percentage effects on adult earnings from adult outcomes come from Heckman, Moon, et al. (2010), Table 3, p. 119. We simply added up gross earnings effects, undiscounted, for all ages up to age 65, for both males and females, and computed percentage gain for treatment group versus controls. Heckman et al. interpolate earnings effects up to age 40, and extrapolate from observed earnings after age 40. 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.

Table 3.1 Comparison of Average Test Percentile and Covariates for TPS Pre-K Entrants and Alumni

Variable	Pre-k entrants			Pre-k alumni			Diff.	<i>p</i>
	M	SE	N	M	SE	N		
Average test percentile	27.031	0.859	1,418	41.897	1.327	1,256	-14.866	0.000
Female	0.499	0.026	1,418	0.482	0.029	1,256	0.017	0.660
Race								
Black	0.325	0.025	1,393	0.345	0.027	1,252	-0.021	0.566
White	0.349	0.025	1,393	0.340	0.027	1,252	0.009	0.810
Hispanic	0.210	0.021	1,393	0.220	0.023	1,252	-0.010	0.752
Native American	0.102	0.016	1,393	0.082	0.017	1,252	0.020	0.388
Asian	0.014	0.006	1,393	0.012	0.007	1,252	0.002	0.862
Lunch status								
Free	0.628	0.025	1,403	0.677	0.028	1,254	-0.049	0.187
Reduced-price	0.146	0.018	1,403	0.102	0.019	1,254	0.045	0.085
Full-price	0.226	0.023	1,403	0.221	0.025	1,254	0.005	0.884
Mother's education								
No high school	0.189	0.023	1,070	0.223	0.027	887	-0.034	0.333
High school	0.262	0.027	1,070	0.263	0.030	887	-0.001	0.986
Some college	0.397	0.029	1,070	0.370	0.034	887	0.027	0.545
College degree	0.152	0.020	1,070	0.144	0.024	887	0.008	0.795
Lives with father	0.622	0.028	1,152	0.589	0.031	994	0.033	0.433
Internet access	0.469	0.028	1,164	0.491	0.032	1,002	-0.021	0.619

NOTE: This table compares regression-based estimates of the values of different variables at the birthday cutoff for TPS program participation in the 2006–07 academic year (September 1, 2001). Thus, 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, based on the child's age.

Table 3.2 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)	15.467*** (1.465)	16.258*** (1.753)	15.723*** (4.416)	12.527*** (3.318)
Age (days)	0.048*** (0.004)	0.039*** (0.004)	0.050*** (0.010)	0.073*** (0.010)
Age*Born before cutoff	0.004 (0.007)	0.015* (0.008)	0.003 (0.019)	-0.027* (0.015)
Black	-5.259*** (0.937)	-4.875*** (1.151)	-6.144** (2.671)	-6.578*** (2.102)
Hispanic	-6.916*** (1.107)	-7.902*** (1.290)	-7.875** (3.045)	0.020 (3.116)
Native American	-0.847 (1.221)	-1.244 (1.583)	-2.047 (2.823)	0.484 (2.623)
Asian	-0.612 (2.706)	-3.234 (3.898)	-1.032 (7.567)	1.458 (4.033)
Female	5.989*** (0.700)	5.525*** (0.833)	5.903*** (1.998)	6.873*** (1.600)
High school	2.789*** (1.060)	2.271** (1.147)	3.837 (3.343)	8.797* (4.584)
Some college	6.489*** (1.186)	6.307*** (1.389)	6.549** (3.208)	12.400*** (4.188)
College or higher	14.169*** (1.780)	9.475*** (3.221)	9.488** (4.503)	22.420*** (4.438)
Lives with father	1.317 (0.829)	0.644 (0.965)	0.685 (2.636)	4.393** (1.803)
Internet access	4.885*** (0.794)	3.855*** (0.965)	4.965** (2.155)	8.297*** (2.048)
Free lunch	-6.379*** (1.049)			
Reduced-price lunch	-5.378*** (1.287)			
Constant	23.592*** (1.780)	17.354*** (1.724)	19.362*** (4.458)	15.149*** (4.651)
Observations	2,657	1,650	343	664
R-squared	0.56	0.54	0.54	0.55
Effect size ^a	0.89	1.21	0.98	0.57
Percentage effect ^b	83.3	112.8	82.5	44.1

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 are in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

^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 the treatment effect divided by the mean of the outcome for the comparison group (pre-K entrants) in the relevant sample.

Table 3.3 Effects of TPS Pre-K Participation on Average Test Score Percentile, 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
Treatment effect	18.090*** (2.000)	20.332*** (5.542)	16.380*** (5.182)	11.743*** (3.664)	8.000 (7.912)	11.008** (4.494)
Effect size ^a	1.37	1.33	0.71	0.82	0.44	0.54
Percentage effects ^b	130.24	112.34	53.72	70.82	36.78	41.76
Observations	1,292	227	297	358	116	367
R-squared	0.56	0.62	0.48	0.51	0.44	0.60

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 are in parentheses. All regressions include the full set of control variables reported in Table 3.2; full results are available on request. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

^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 the treatment effect divided by the mean of the outcome for the comparison group (pre-K entrants) in the relevant sample.

Table 4.1 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	\$27,179	\$30,548	\$24,610	\$17,643	\$12,020	\$16,539
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	5.4	6.8	3.5	3.6
Program costs	\$8,806	\$8,806	\$8,806	\$4,403	\$4,403	\$4,403
Ratio of program earnings benefits to costs	3.09	3.47	2.79	4.01	2.73	3.76
Ratio of incremental earnings benefits from full-day over half-day, compared to incremental costs	2.17	4.21	1.83			

NOTE: The present value of adult earnings increases are derived, using techniques discussed in the text, from the treatment effects on kindergarten test scores reported in Table 3.3. See text for derivation of baseline predicted future earnings figures and program cost figures. 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 incremental earnings benefits-to-costs ratios for full-day versus half-day is the difference in earnings benefits divided by the differences in costs.