Running Head: EFFECTS OF PRE-K

The Effects of Universal Pre-K on Cognitive Development

William T. Gormley, Jr. and Ted Gayer

Public Policy Institute

Georgetown University

Deborah Phillips

Department of Psychology

Georgetown University

Brittany Dawson

Center for Research on Children in the U.S.

Georgetown University

Washington, D.C.

Abstract

Many studies of the effects of pre-kindergarten (pre-K) programs suffer from selection bias because children who did not attend the programs (the control group) likely differ in both observable and unobservable ways from those who did attend (the treatment group). In this study of Oklahoma's universal pre-K program, we rely on a strict birthday eligibility criterion to compare "young" kindergarten children who just completed pre-K to "old" pre-K children just beginning pre-K. This "regression discontinuity" design reduces the threat of selection bias. Our sample consists of 1,567 pre-K children and 1,461 kindergarten children who just completed pre-K. We estimate the impact of the pre-K treatment on Woodcock-Johnson Achievement test scores. We find test impacts of 3.00 points (0.79 of the standard deviation for the control group) for the letter-word identification score; 1.86 points (0.64 of the standard deviation of the control group) for the spelling score; and 1.94 points (0.38 of the standard deviation of the control group) for the applied problems score. Hispanic, black, white, and Native American children all benefit from the program, as do children in diverse income brackets, as measured by school lunch eligibility status. Although we cannot generalize about the relative merits of full-day and half-day programs, we find benefits for both types of programs, and for both programs for three of four racial/ethnic groups. We conclude that Oklahoma's universal pre-K program has succeeded in enhancing the school readiness of a diverse group of children.

The Effects of Universal Pre-K on Cognitive Development

Enrollment in a state-funded pre-kindergarten (pre-K) program is becoming a common pathway into kindergarten for preschoolers in the U.S. (Pianta & Rimm-Kaufmann, in press). The number of states that administer publicly funded pre-K services has soared from 10 in 1980 to 38 in 2002, with combined enrollments exceeding 700,000 children and total state spending exceeding $2.5 billion (Barnett, 2004; Gilliam and Zigler, 2004). Propelled by national school readiness goals, these programs have as their central aim promoting the acquisition of skills, knowledge, and behaviors that are associated with success in elementary school.

Most state pre-K programs are targeted to disadvantaged children, but six states have established programs that might be described as “universal” – Florida, Georgia, Massachusetts, New York, Oklahoma, and West Virginia. In practice, a universal program means that the program is universally available (or nearly so) but that parents are free to enroll their children or not as they see fit. The existing universal programs – some mature (Georgia, New York, Oklahoma), some just beginning (Florida, Massachusetts, West Virginia) -- are aimed at four-year-olds. The District of Columbia has a relatively mature universal program, for four-year-olds. Los Angeles County has committed itself to such a program, for four-year-olds and three-year-olds as well.

This paper reports on the school readiness of children who attended the universal pre-K program in Tulsa, Oklahoma during the 2002-03 school year. Using a quasi-experimental regression-discontinuity design that reduces the threat of selection bias, we estimate the overall effects of exposure to pre-K, also for children varying in race, ethnicity and income, and for children in full-day and half-day programs.

What Can We Expect From Pre-kindergarten Programs?

There are reasons to expect both promising and disappointing results from research on the developmental consequences of attending a pre-K program. A series of well-designed and implemented model preschool programs have shown significant short-term and some long-term effects on young children's cognitive growth. Such effects have been reported for small demonstration programs such as the Perry Preschool Project (Schweinhart, Barnes, Weikart, Barnett & Epstein, 1993) and for carefully controlled early interventions such as the Abecedarian program (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002) and the Infant Health and Development Project (McCarton, Brooks-Gunn, Wallace, & Bauer, 1997). Evidence on Head Start remains controversial, although carefully designed studies have documented positive effects on children's early learning (see Currie & Thomas, 1995; Garces, Thomas & Currie, 2002; Abbott-Shim, Lambert, & McCarty, 2003), as has a recent random-assignment evaluation of the Early Head Start program (Love, in press; Love, Kisker, Ross, Brooks-Gunn, et al., 2002).

Nevertheless, in a review of the long-term academic impacts of both model and large-scale public preschool programs, including Head Start, Barnett (1995, 1998) found that public programs often had weaker effects than the generally higher-quality and better implemented model programs. Moreover, research on the developmental implications of more naturally-occurring child care experiences has generated mixed evidence. Links between higher quality child care environments and children's cognitive and language development have been repeatedly documented (NICHD ECCRN, 2000; Peisner-Feinberg, Burchinal, Clifford, Culkin, Howes & Kagan, 2001), but they are often weak and diminish when extensive selection controls are included in estimation models (NICHD ECCRN, 2003). They are found most consistently for children who, post-infancy, were enrolled in center-based arrangements (NICHD ECCRN, 2000) and appear to be stronger for children growing up in low-income households who typically receive less support for cognitive and language development (Loeb, Fuller, Kagan, & Carrol, 2004; Votruba-Drzal, Coley, & Chase-Lansdale, 2004).

It is difficult to locate state pre-K programs along this spectrum of early childhood options from very high-quality model early intervention programs to highly variable community-based child care. Prior research suggests that the typical school-based preschool program is of higher quality than the typical child care program serving low-income children (Goodson & Moss, 1992; Phillips, Voran, Kisker, Howes, & Whiteboook, 1994). Nevertheless, emerging descriptive data indicate that pre-K programs, like child care, are characterized by extensive variation. For example, state-funded pre-K programs range from as short as 2.5 hours per day to as long as 10 hours per day (Bryant et al., 2004). A handful of states, including Oklahoma, require that all pre-K teachers have a college degree and certification in early childhood education, while many others require only a Child Development Associate (CDA) certificate. This variability confounds efforts to relate findings that are emerging from evaluations of pre-K programs to prior evidence from either model intervention or more typical early childhood and child care programs. It also may have important public policy implications, because, for example, pre-K children living in poverty are more likely to be enrolled in a program staffed by teachers with lower qualifications than children with greater resources (Clifford et al., 2003).

Available Research on Pre-Kindergarten Programs

Preliminary results from a growing body of research on the effects of pre-K programs are encouraging, but not entirely convincing. A careful meta-analysis of state-funded preschool programs in 13 states found statistically significant positive impacts on some aspect of child development (cognitive, language, or social) in all of the states. A study of Georgia's universal pre-K program found that 82% of former pre-K students rated average or better on third-grade readiness in comparison to national norms (Henry et al., 2001). A more recent study found that economically disadvantaged children attending Georgia's pre-K program began preschool scoring below national norms on a letter and word recognition test but began kindergarten scoring above national norms (Henry et al., 2003). Similarly, a recent Michigan study, using a non-experimental research design, reached positive conclusions: in kindergarten, teachers rated students who attended a pre-K program higher in language, literacy, math, music, and social relations; students who attended a pre-K program were more likely to pass the Michigan Educational Assessment Program's reading and mathematics tests (Xiang and Schweinhart, 2002). An analysis of the ECLS-K data found that kindergarten students who had attended a pre-K program scored higher on reading and math tests than children receiving parental care. Kindergarten students who had attended a preschool or child care center also experienced improvements, but the pre-K participants' improvements were more substantial (Magnuson et al., 2003).[1] Finally, evidence from a prior evaluation of Tulsa, Oklahoma's pre-K program using a locally-developed test indicated that Hispanic and African-American children, but not whites, benefited significantly in cognitive and language domains (Gormley & Gayer, 2005; Gormley & Phillips, 2005).

In many instances, these child outcome findings are not disaggregated by characteristics of the child or the pre-K setting. Some pre-K evidence, however, as well as the early intervention literature, suggests that the largest effects of such programs accrue to children from lower-income families and from non-white racial groups (Gormley and Gayer, 2005; Campbell et al., 2002; Reynolds et al., 2001). The Early Head Start evaluation, for example, found that parenting and social-emotional developmental outcomes were strongest for African-Americans, and that families experiencing three risk factors (e.g., single parenthood, receiving public assistance, teen parenthood) showed stronger effects than families with either fewer or more risk factors (Love et al., 2002). Within the early intervention literature, stronger results have also been reported for more intensive programs measured as hours of contact, part-day versus full-day, total years of intervention, and extent of compliance with program standards (Campbell & Carmey, 1994; Love et al., 2002; Ramey et al., 1992; Reynolds et al., 2001; DeSiato 2004). For these reasons, we examine differential pre-K effects by family income (using free lunch eligibility as a proxy for family income) and racial-ethnic group of the children and by their enrollment in half- or full-day programs.

Methodological Shortcomings of Previous Research

Unfortunately, the vast majority of research on this important issue falls short of scientific standards for drawing causal inferences (Gilliam & Ripple, 2004; Gilliam & Zigler, 2001, 2004), though most researchers do make a good-faith effort to control for relevant variables. Virtually all published evaluations of state pre-K programs, as well as the national studies, fail to correct for selection bias, many rely on tests that have not been normed or validated, and it is not uncommon for studies to rely on pre-K teacher reports in pretest-posttest designs, thus introducing strong evaluator bias. None of the studies examined by Gilliam and Zigler (2001, 2004) used random assignment and only the Oklahoma study and one other study used a comparison group that credibly addressed the selection bias problem. These methodological shortcomings are serious, because children who select into preschool programs often differ in terms of their observable characteristics (e.g., social class) from those who do not. If they differ in terms of observable characteristics, they probably differ in terms of unobservable characteristics as well, thus potentially introducing omitted variable bias. The current study improves upon the extant literature by adopting a design that reduces the threat of selection bias. It also uses a standardized assessment instrument and relies upon college-educated and specially trained teachers to administer the tests to the children.

Research Questions

The current study was motivated by three central research aims. First, we examine the overall effect of the Oklahoma universal pre-K program, using Tulsa as the study site, on children's school readiness as assessed by three subtests of the Woodcock-Johnson Achievement Test. We compare these results to estimates obtained using a traditional cross-sectional analysis that does not address the selection problem. Second, we measure differences in program impact by disaggregating the results for children who vary in their race/ethnicity and family income (as measured by eligibility for free or reduced price school lunch). Third, we further disaggregate our results by comparing the performance of children enrolled in full-day and half-day programs, by racial/ethnic group.

Method

Why Oklahoma?

Oklahoma established a universal pre-K program, for four-year-old children, in 1998, after having administered a targeted program aimed at economically disadvantaged children for eight years. Under the 1998 legislation, participating school districts receive state aid for every four-year-old they enroll in a pre-K program. Each of the state's 543 public school districts is free to participate or not. As of 2002-03, 91 percent of the state's school districts were participating. Each school district is free to run half-day or full-day programs or some combination of the two. As of 2002-03, approximately 56 percent of the state's pre-K programs were half-day, 44 percent full-day.[2] All teachers in the program are required to have a B.A. degree and an early childhood certificate. Group sizes are a maximum of 20 with a required ratio of ten students per teacher. As a result, most classrooms also have an assistant teacher, who has no specific education or training requirements.

Oklahoma's universal pre-K program was a particularly inviting target of opportunity because of its strong commitment to both universality and quality. Unlike New York's program, which has a penetration rate of approximately 33 percent, Oklahoma has a penetration rate of approximately 63 percent.[3] Unlike Georgia's program, which does not require all teachers to have a college degree, Oklahoma requires its teachers to have a college degree and to be early childhood certified. Furthermore, Oklahoma pays its pre-K teachers at precisely the same rates as other elementary and secondary school teachers. Oklahoma's program reaches more four-year-olds than any other program in the nation, and its quality standards are unsurpassed by any other program.

Within Oklahoma, Tulsa was an attractive research site for several reasons. First, it is the largest school district in Oklahoma, with 41,048 students in 2002-03.[4] Second, it has a diverse student population, in terms of race and ethnicity. In 2002-03, the student body was 41 percent white, 36 percent black, 13 percent Hispanic, 9 percent Native American, and 1 percent Asian.[5] Third, it administers tests to both four-year-olds and five-year-olds at the same point in time, at the beginning of the school year. As discussed more fully below, this is critical to our research design. Fourth, the Superintendent of the Tulsa Public Schools (TPS) granted us permission to administer the three subscales of the Woodcock-Johnson Achievement Test (WJ-ACH-III) in addition to their homegrown test.

Participants

Participants consisted of pre-K and kindergarten children enrolled in the Tulsa, Oklahoma public schools. Of 1,845 pre-K students, 85.0% (1,567) were tested; of 3,727 kindergarten students, 84.5% (3,149) were tested.[6] In general, the racial/ethnic characteristics and gender of tested children closely resemble the universe of children from which they were drawn. As Table 1 indicates, however, there are some discrepancies. Among pre-K students, blacks were more likely to be tested; among kindergarten students, Hispanics and those eligible for a free lunch were less likely to be tested and those eligible for a full price lunch were more likely to be tested.

Because both kindergarten and pre-K children took the same three tests at the beginning of the school year, in effect we have test data for a treatment group and a control group that both selected into the Tulsa pre-K program. Our treatment group consists of the kindergarten children who were enrolled in the Tulsa pre-K program the previous year. At the time of testing, these children had just completed the treatment (i.e., Tulsa pre-K). The control group consists of children who at the time of testing had just begun Tulsa pre-K (and had thus selected into the same program). This is critical to our analysis strategy, described below.

It is important to emphasize that our treatment group consists of children who chose the treatment the previous year. We do not have as much information as we would like on the kindergarten children who did not participate in the Tulsa pre-K program. They could have opted for a private pre-K program, a day care center, a Head Start program, or no program. Thus, our research design only estimates the treatment-on-the-treated effect, which is the impact on test scores of attending Tulsa pre-K. We cannot estimate the intent-to-treat effect, which is the impact on the population's test scores of making the Tulsa pre-K program available to everyone.

Measures

The test instrument consisted of three subtests of the Woodcock-Johnson Achievement Test (WJ-ACH-III). The Woodcock-Johnson Achievement Test is a nationally normed test that has been widely used in studies of early education and its consequences, including studies with racially and socio-economically mixed samples. For example, it has been used in the Head Start Family and Child Experiences Survey (FACES) and in the Abecedarian study. It has also been used in studies of maternal employment on children's educational outcomes (Chase-Lansdale et al., 2003). We specifically chose subtests that are thought to be appropriate for relatively young children, including pre-K children (Mather &Woodcock, 2001). They are: Letter-Word Identification, Spelling, and Applied Problems.

The Letter-Word Identification Subtest measures pre-reading and reading skills. It requires children to identify letters that appear in large type and to pronounce words correctly (the child is not required to know the meaning of any particular word). The Spelling Subtest measures pre-writing and spelling skills. It measures skills such as drawing lines and tracing letters and requires the child to produce uppercase and lowercase letters and to spell words correctly. The Applied Problems Subtest measures early math reasoning and problem-solving abilities. It requires the child to analyze and solve math problems, performing relatively simple calculations.

Procedure

The tests were administered by teachers in the TPS kindergarten and pre-K programs. Because teachers administered the test at the beginning of the school year, they were in effect gathering data that would be used to assess someone else's performance, rather than their own. Teachers were trained to administer the tests at one of two training sessions held in Tulsa in August 2003.[7] Teachers received modest compensation for attending a training session. All TPS teachers are college educated.

The testing took place, for the most part, during the first week of school (September 2-8, 2003). Under a special waiver from the state Department of Education, TPS is permitted to designate the first week of school as a testing period. A small number of TPS schools, operating on what is called a "continuous learning" calendar, began school three weeks ahead of the rest and conducted their tests earlier than other programs.

As discussed below, TPS classrooms include a substantial number of Hispanic children, many of whom come from Spanish-speaking households. Teachers were instructed to administer the test exclusively in English. Teachers were also instructed to administer the test to all children, unless it proved impossible to get any meaningful response.

Data Analysis

Because the Oklahoma pre-K program is voluntary, comparing test scores of kindergarten children who completed the program to kindergarten children who did not is likely to suffer from selection bias. Certain families are more likely to select into the pre-K program, and these same families will have unobservable characteristics that influence test scores. Thus a traditional comparison of kindergarten children exposed to pre-K to kindergarten children not exposed to pre-K, even using controls for selection bias, could lead to spurious results. This bias can be in either direction. If families with unobservable characteristics that contribute to under-performing children are more likely to select into Tulsa's pre-K program, then test impacts will be underestimated. If families with unobservable characteristics that contribute to over-performing children are more likely to select into the program, then test impacts will be overestimated.

Table 2 presents evidence of the likelihood of selection bias in a traditional cross-sectional analysis. The top three rows show that mean test scores are indeed significantly higher for the kindergarten children who attended Tulsa pre-K relative to those who did not. However, the treatment and control groups differ in many observable ways. The children who attended Tulsa pre-K are less likely to be on full price lunch, more likely to be on reduced price lunch, and they are more likely to be non-white (though less likely to be Hispanic). They are also less likely to have mothers who did not finish high school and more likely to have mothers who completed some college education. These differences suggest that the cross-sectional analysis might result in biased estimates of the true impacts of the Tulsa pre-K program.

Comparison of control and treatment groups. An alternative approach is possible because both kindergarten and pre-K children took the same three tests at the beginning of the school year, as described above. We have a treatment group consisting of the kindergarten children who were enrolled in Tulsa pre-K the previous year and were tested just after completing the treatment. The control group consists of children who at the time of testing had just begun Tulsa pre-K (and had thus selected into the same program). Comparing these two groups should reduce the selection problem.

A potential problem with this strategy is that, while the selection criteria may be the same across the two years, the control and treatment groups may still have different characteristics. For example, if socio-demographic characteristics are changing over time within Tulsa, or if the selection decision changes over time, then this strategy would not lead to similar control characteristics across the treatment and control groups.

The first set of columns in Table 3 confirms this concern. This set of columns compares the test scores and the observable characteristics of the treatment group (the children who just finished Tulsa pre-K) to the control group (the children who are just beginning Tulsa pre-K). While there is a statistically significant difference in test scores, there are also many differences in the observable characteristics, which could suggest biased estimates of test score differences. The treatment group has a smaller proportion of Hispanics and mothers who did not complete high school. The treatment group has a higher proportion of Native Americans, mothers with only a high school degree, and mothers with a college degree.

Whether a child is part of the treatment group or control group depends on the child's date of birth. Under Oklahoma law, children qualified to attend TPS pre-K in academic year 2002-03 if, and only if, they were born on or before September 1, 1998 (and after September 1, 1997). If a child missed this cut-off date, then he or she would have to wait another year to enroll in the TPS pre-K program. According to TPS officials, exceptions to the admissions policy are extremely rare. This strict birthday requirement allows us to restrict our comparison of means to children near the cut-off date. That is, we can compare the mean test scores of children in our treatment group who just made the age cut-off qualification to the mean test scores of children in our control group who just missed making the age cut-off the previous year.

In the second and third sets of columns of Table 3, we narrow the range of observations to those children born closer to the cut-off birthday. We see that this narrowing results in a decrease in the test score differences, which is likely attributable to a reduced influence of age. However, we also see that some of the differences in observable characteristics disappear as we narrow the range around the cut-off date. In other words, the closer that we focus our comparison of means around the cut-off birthday, the better we are able to control for confounding influences on test scores. This helps us to isolate the treatment effect of TPS pre-K.

Regression-discontinuity design. The strict birthday cut-off qualification provides us with a regression-discontinuity research design, in which assignment to treatment is based only on the cut-off score, which in this case is age (see Cook and Campbell, 1979). We can estimate the relationship between age and test scores for the children who qualified for TPS pre-K the previous year (our treatment group), and we can estimate the relationship between age and test scores for the children who did not qualify for TPS pre-K the previous year (our control group). These two regression lines then allow us to compare the expected test score of a (treated) child who just made the cut-off to the expected test score of a (control) child who just missed the cut-off qualification. Essentially, we use the full data set of children on both sides of the cut-off qualification to estimate whether (and to what degree) there is a test-score jump at the limit where the range around the cut-off is infinitesimally small (see Gormley and Gayer, 2005, for a fuller description of this analytical framework). The fundamental premise supporting this regression-discontinuity research design is that a child who just made the cut-off date and a child who just missed the cut-off date should have statistically similar characteristics, except that the former child has received the treatment whereas the latter has not.

Figure 1 provides a hypothetical illustration of this regression-discontinuity research design. The broken line to the right of the cut-off date shows the hypothetical test scores of the treatment group, and the solid line to the left of the cut-off date shows the hypothetical test scores of the control group. The key challenge in estimating the effect of TPS pre-K is to estimate the counterfactual, or test score outcomes for treated children had they not been treated. The dotted line to the right of the cut-off date depicts the counterfactual. The regression-discontinuity design assumes that the counterfactual is continuous at the cut-off date, so any estimated increase in test scores for the treated children relative to the counterfactual is attributed to the TPS pre-K program. Our estimation strategy effectively compares the children who just barely missed the cut-off to those who just barely made the cut-off.[8] These estimates are derived using the full data set, which contains observations ranging from children born 12 months before the cut-off to children born 12 months after the cut-off.

The key identifying assumption is that the unobservable characteristics of the children born immediately before the cut-off date do not differ from the unobservable characteristics of the children born immediately after the cut-off date. Because the data points close to the cut-off date are the key to identification, restricting the data set to ranges narrower than the 12-month margin should reduce any potential bias; however, the loss of data points would also reduce the precision of the estimates. We rely primarily on the results using the 12-month margin, yet we also report results using data restricted to 6-month and 3-month margins.

The goal of the regression-discontinuity design is to estimate predicted test scores at the cut-off birthday, approaching from both sides of the cut-off. We accomplish this by regressing the test scores on a second-order polynomial of the difference between the birthday and cut-off date. We conduct this regression separately for the sample of children who missed the cut-off date (control) and the sample of children who made the cut-off date (treatment). We chose the second-order, or quadratic, polynomial functional form because it is more flexible than a linear relationship. However, the results are fairly robust across specifications.

The last set of columns in Table 3 shows the differences in test scores (discussed below) and in observable characteristics at the cut-off birthday using the quadratic specification. It suggests that our regression-discontinuity design alleviates (though perhaps does not eliminate) the selection bias. Most of the mean observable characteristics are the same at the discontinuity, but the percent Hispanic and the percent with mothers with no high school degree are still higher for the control group.

Instead of estimating equations on both sides of the cut-off birthday and computing the predicted difference at the cut-off, we can also estimate the same impact in a single-equation model in which the dependent variable is the test score, and the independent variables include the difference in days between the birthday and the cut-off day, the square of this term, a cut-off dummy variable, and interaction variables. This single equation with interaction terms is equivalent to computing two different equations; however, it is more practical to implement. We use this single-equation approach in our reported results, and we also include other observable covariates in order to estimate the effects of these characteristics on test outcomes. These other covariates measure whether the child is on no free lunch, partial free lunch, or full free lunch, the child's mother's education (no high school degree, high school degree only, some college experience, and college degree), the race/ethnicity of the child (white, black, Hispanic, Native American, or Asian), and the gender of the child.

Results

Positive Effects for the Total Sample

The last set of columns of Table 3 shows the predicted test impacts using the quadratic regression-discontinuity method. The results suggest that the letter-word identification scores increase by 3.05 points (0.80 of the standard deviation for the control group), the spelling scores increase by 1.88 points (0.65 of the standard deviation for the control group), and the applied problems scores increase by 1.96 (0.38 of the standard deviation for the control group).

In the first set of columns on Table 4, we replicate this estimation in a single-equation framework, except that we include additional covariates. We add these covariates to estimate the impact they have on test scores. Our research design should credibly replicate an experimental design if the characteristics that influence test scores do not vary discontinuously at the cut-off date. In such a case, adding the covariates should not significantly change the results. The results in the first set of columns in Table 4 indeed closely match the quadratic results in Table 3. We find test impacts of 3.00 points (0.79 of the standard deviation for the control group) for the letter-word identification score; 1.86 points (0.64 of the standard deviation of the control group) for the spelling score; and 1.94 points (0.38 of the standard deviation of the control group) for the applied problems score. We should also note that the coefficients for the quadratic term and interaction variables in Table 4 are mostly not statistically different from zero. This suggests that the relationship between age and test scores is primarily a linear one, and that the effect of the TPS pre-K treatment does not vary by age.

In order to compute the proportionate increase in test scores attributable to TPS pre-K, we estimate the ratio of the treatment effect to the predicted test score of a child born at the cut-off date who did not receive the treatment. We find that the treatment leads to a 52.95% increase in the letter-word score, a 26.42% increase in the spelling score, and a 17.94% increase in the applied problems score.[9]

As discussed earlier, a cross-sectional comparison of kindergarten children who attended Tulsa pre-K to kindergarten children who did not would likely suffer from selection bias. Though not reported in the tables, we conducted such a naïve regression and found test impacts of 2.79 for the letter-word identification test, 1.46 for the spelling test, and 1.56 for the applied problems test. These naïve cross-sectional results understate the test impact compared to our regression-discontinuity results.

The other sets of columns in Table 4 replicate the results using narrower margins around the cut-off birthday. Identification in the regression-discontinuity research design comes from the observations in the neighborhood around the cut-off birthday, so narrowing the margin should reduce any bias. However, because narrowing the margin substantially reduces the sample size, the standard errors increase. Given the precipitous drop in the number of observations as the margin is narrowed, the 12-month margin estimates are our preferred estimates of the TPS pre-K impact. If one looks at all the columns together, the results mostly indicate that treatment effects remain statistically different from zero; however, the point estimates do vary somewhat in their magnitudes.

Subgroup Results: Race/Ethnicity and Free Lunch Eligibility

In Table 5 we report results separately by race/ethnicity, by free lunch eligibility (a surrogate for income), by full-day versus half-day program, and for race/ethnicity by full-day versus half-day program. The latter is important because black children are more likely than other children to be enrolled in a full-day program.

We find sizable test improvements across all racial/ethnic groups. For Hispanic children, the letter-word identification scores increase by 4.15 (1.50 of the standard deviation for the control group), the spelling scores increase by 2.66 points (0.98 of the standard deviation for the control group), and the applied problems scores increase by 4.97 points (0.99 of the standard deviation for the control group). For black children, the letter-word identification scores increase by 2.91 (0.74 of the standard deviation for the control group), the spelling scores increase by 1.47 points (0.52 of the standard deviation for the control group), and the applied problems scores increase by 1.68 points (0.38 of the standard deviation for the control group). For Native American children, the letter-word identification scores increase by 3.56 (0.89 of the standard deviation for the control group), the spelling scores increase by 2.24 points (0.72 of the standard deviation for the control group), and the applied problems scores increase by 3.08 points (0.60 of the standard deviation for the control group). For white children, the letter-word identification scores increase by 3.02 (0.76 of the standard deviation for the control group), the spelling scores increase by 2.07 points (0.72 of the standard deviation for the control group), and the applied problems scores do not have a statistically significant increase.

We also find test impacts no matter what the free lunch eligibility status. For children receiving full price lunch, the letter-word identification scores increase by 2.69 (0.63 of the standard deviation for the control group), the spelling scores increase by 1.59 points (0.54 of the standard deviation for the control group), and the applied problems scores increase by 1.54 points (0.29 of the standard deviation for the control group). For children receiving reduced price lunch, the letter-word identification scores increase by 5.00 (1.04 of the standard deviation for the control group), the spelling scores increase by 2.81 points (0.97 of the standard deviation for the control group), and the applied problems scores do not have a statistically significant increase. For children receiving full free lunch, the letter-word identification scores increase by 2.79 (0.81 of the standard deviation for the control group), the spelling scores increase by 1.75 points (0.65 of the standard deviation for the control group), and the applied problems scores increase by 2.10 points (0.45 of the standard deviation for the control group).

For both full-day and half-day programs, we find positive and statistically significant impacts for all three tests. For race/ethnicity by full-day versus half-day program, we find positive point estimates for almost all of the test impacts. In some instances, due to small sample sizes, standard errors are so high that statistically significant findings are unlikely. For example, we find high point estimates for the test impacts for Native American children in a full-day program, but the very small sample size substantially increases the standard errors for these estimates. With this exception, we find strong test impacts for all racial and ethnic groups across full-day and half-day programs.

Discussion

This study examines the effects of school-based universal pre-K attendance on children at the point of kindergarten entry. The results provide solid support for the benefits that such a program can have on the test scores of young children of differing ethnic and racial groups and from differing socio-economic backgrounds. Specifically, for those who select into the pre-K program, the program was found to have statistically significant effects on children’s performance on cognitive tests of pre-reading and reading skills, pre-writing and spelling skills, and math reasoning and problem-solving abilities.

The program appears to have its largest effects on the letter-word identification subtest, which assesses pre-reading abilities, followed by those on spelling, and finally applied problems. The relatively greater gains in pre-reading skills may reflect the Tulsa Reads program, which began in the fall of 2001. Under this program, pre-K teachers (and other teachers) received about seven days of professional development aimed at helping them to teach students how to read. The professional development days, distributed throughout the 2001-02 school year, were dropped in 2002-03, primarily because of acute budget difficulties. However, the intensive training, provided just a year earlier, may have given teachers enough momentum to sustain reading activities for the children who attended pre-K in this study. In contrast, the school district's pre-numeracy campaign, Tulsa Counts, did not begin until the fall of 2003, just after our testing took place.

In this study, we have reported effect sizes of 0.79 of a standard deviation for letter-word identification, 0.64 of a standard deviation for spelling, and 0.38 of a standard deviation for applied problems. These effect sizes exceed those reported for other state-funded pre-K programs, which range from 0.23-0.53 (Gilliam and Zigler 2001),[10] and for pre-K programs generally, which range from 0.10 to 0.13 (Magnuson et al. 2004: 5). They also exceed those reported for high-quality child care programs, which seldom exceed 0.10 (NICHD and Duncan 2003; Peisner-Feinberg et al. 2001) . The Abecedarian project, widely acknowledged as a highly successful early intervention program, reported effect sizes of 0.73 and 0.79 for children ages four and five (Ramey et al. 2000), and the highly-praised Perry Preschool program reported effect sizes of 0.60 (Ramey, Bryant, and Suerez 1985). In short, the effect sizes reported here fall somewhere in between those of average state-funded pre-K programs and the very best early intervention programs; they substantially exceed those of high-quality child care programs.

Positive Effects for Diverse Groups of Children

The pre-K program was found to benefit children from all racial/ethnic groups that comprise substantial portions of the Tulsa population: Hispanics, blacks, Native

Americans, and whites. This important finding differs somewhat and adds to prior results regarding the Oklahoma programs (Gormley & Gayer, 2005; Gormley & Phillips, 2005). Specifically, based on a locally-designed testing instrument, Hispanics and blacks but not whites were found to benefit from the 2001-02 pre-K program. We noted that there were not a sufficient number of Native Americans for analysis and we cautioned that the absence of positive findings for white children could be due to “ceiling effects” associated with the homegrown testing instrument, which included a fixed menu of 26 items. This year’s findings suggest that Native American and white children also benefit from the pre-K program. The Woodcock-Johnson Achievement Test, with its highly expandable set of exam questions, appears to have been better equipped to capture positive effects for white children, who were more likely to score at the high end of the 26-item test than other children. The presence of statistically significant positive findings for Native Americans enrolled in TPS pre-K in 2002-03, as opposed to the previous study, is probably attributable to a larger sample size.

The pre-K program was also found to benefit children from diverse income brackets, including children eligible for a full price lunch, a reduced price lunch, and no lunch subsidy at all. This finding also differs from the prior evaluation, in which strong positive effects were reported for free lunch children, limited positive effects for reduced price lunch children, and no positive effects for full price lunch children (Gormley and Gayer, 2005). Again, the current reliance on the standardized and well-validated Woodcock-Johnson Achievement test may explain the difference, particularly in its capacity to capture program impacts for more advantaged children (full price lunch children).

Effects for Full Versus Half Day Programs

It is important to note that one cannot compare the estimated test impacts across sub-groups in Table 5. For example, the greater estimated test impacts for Hispanic children relative to black children does not necessarily imply that a representative Hispanic child will gain more from the program than a representative black child. The Hispanic results measure the test impacts for Hispanic children who chose to select into the program. Likewise, the results for black children measure the test impacts for black children who chose to select into the program.

Similarly, we cannot compare the relative merits of full-day and half-day programs because different types of children select into full-day pre-K versus half-day pre-K. In general, black children are more likely to select into a full-day pre-K, while white children are more likely to select into a half-day pre-K. However, we can say that children in half-day and full-day programs both experience benefits from pre-K and that this is true for three of four racial/ethnic groups. Although we find no statistically significant effects for Native Americans enrolled in full-day programs, we caution that this subgroup is the smallest of all, which could account for the absence of statistically significant findings.

Practical Significance of the Findings

While statistically significant, it is important to ask if these findings are substantively significant as well. One way to answer this question is to convert raw test scores into age-equivalent test scores. If we look first at letter-word identification scores, the average raw score for children not yet exposed to pre-K at the regression-discontinuity point (i.e., the "old" pre-K children) is 5.66, which corresponds to an age equivalent score of 4-7 (or 4 years, 7 months). The average raw score for children exposed to pre-K at the regression-discontinuity point (i.e., the "young" kindergarten children) is 8.70, which corresponds to an age equivalent score of 5-2 or 5-3 (5 years, 2 or 3 months). If we perform the same calculations for spelling, we find that children not yet exposed to pre-K have an age equivalent score of 4-6, while children exposed to pre-K have an age-equivalent score of 5-0 or 5-1. As for applied problems, children not yet exposed to pre-K have an age equivalent score of 4-5, while children exposed to pre-K have an age-equivalent score of 4-9. In short, children exposed to pre-K experience a gain of seven to eight months in letter-word identification, six to seven months in spelling, and four months in applied problems, above and beyond the gains of aging or maturation.

Impact of Controls for Selection Bias

A final result that warrants highlighting concerns the comparison of findings from the naïve regression and the more methodologically rigorous regression-discontinuity results. The naïve regression produced findings that actually underestimated the impacts of the pre-K program relative to the results emerging from the analysis that reduced the threat of selection bias. Prior evaluations of pre-K programs that fail to address selection bias could under or overestimate program impacts, depending on which children select into the program.

Conclusion

Public officials at all levels of government have engaged in spirited debates over the best means of supporting school readiness for young children. Our research supports the proposition that a universal pre-K program financed by state government and implemented by the public schools can improve pre-reading, pre-writing, and pre-numeracy skills for a diverse cross-section of young children. It is important to emphasize, however, that our research design estimates the effect of Tulsa pre-K on those children whose families chose to enroll them in the program. That is, we estimate the impact on test scores of attending Tulsa pre-K; we cannot estimate the impact on the population's test scores of making the Tulsa pre-K program available to everyone.

Can the Oklahoma experience be replicated elsewhere? While it is difficult to generalize, we should note Oklahoma's high teacher education requirements, which other research has found to be a strong predictor of high quality environments for young children (NICHD ECCRN, 1999, 2002). Also noteworthy is Oklahoma's willingness to compensate pre-K teachers at the same level as elementary and secondary school teachers in the public schools, which helps pre-K programs to recruit and retain talented teachers. Without such education requirements and pay levels, other states that opt for universal pre-K might experience weaker results. Other departures from the Oklahoma paradigm could also lead to different outcomes. For example, informal observations within a subset of the Tulsa pre-K classrooms indicated a consistently strong emphasis on academic instruction, albeit through widely differing instructional strategies. Some teachers have created their own curriculum, while others have borrowed from such standardized curricula as Curiosity Corner, the Waterford Early Learning Program, Integrated Thematic Instruction, Creative Curriculum, and Direct Instruction. Differences in children's characteristics could also lead to different outcomes.

Clearly, we need to know more about the mechanisms that have enabled Oklahoma to boost the school readiness of young children in each of the areas we assessed: letter-word identification, spelling, applied problems. This will involve getting inside the classroom doors, as some researchers have (Bryant et al., 2002; Pianta et al., 2003; Stipek and Byler, 2003), to understand better the relationship between pedagogy, other aspects of the classroom activities and climate, and test scores. We also need to examine a broader spectrum of assessments, including socio-emotional and motivational outcomes, and to compare pre-K with other prominent early childhood programs, such as Head Start. A universal pre-K program may or may not be the best path to school readiness. It is, however, a promising path with considerable potential.

REFERENCES

Abbott-Shim, M., Lambert, R. and McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a Head Start program and the program's wait list. *Journal of Education for Students Placed at* Risk 8, 191-214.

Barnett, W.S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children* 5(3), 25-50.

Barnett, W.S. (1998). Long-term effects on cognitive development and school success. In W.S. Barnett & S.S. Boocock, (Eds.), *Early care and education for children in poverty* (pp. 11-44). Albany, NY: State University of New York Press.

Barnett, S., Hustedt, J., Robin, K. & Schulman, K. (2004). *The State of Preschool: 2004 State Preschool Yearbook*. New Brunswick, N.J.: National Institute of Early Education Research.

Bryant, D., Clifford, R., Early , D., Pianta, R., Howes, C., Barbarin, O., & Burchinal, M. (2002, November). *Findings from the NCEDL Multi-State Pre-Kindergarten Study*. Annual meeting of the National Association for the Education of Young Children, New York, NY.

Bryant, D., Cliffford, R., Saluga, G., Pianta, R., Early, D., Barbarin, O., Howes, C., & Burchinal, M. (2004). Diveristy and directions in state pre-kindergarten programs. Unpublished manuscript: University of North Carolina at Chapel Hill.

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

Campbell, F.A. and Carmey, C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, 65: 684-98.

Chase-Lansdale, L., Moffitt, R., Lohman, B., Cherlin, A., Coley, R., Pittman, L., Roff, J., & *Votruba*-Drzal, E. (2003). Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents, *Science*, 299, 1548-1552.

Clifford, D., Barbarin, O., Change, F., Early, D., Bryant, D., Howes, C., Burchinal, M. & Pianta, R. (2003). What is pre-kindergarten? Trends in the development of a public system of pre-kindergarten services. Unpublished manuscript: University of North Carolina at Chapel Hill.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Co.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, 85 (3), 341-364.

DeSiato, Donna (2004). Does prekindergarten experience influence children's subsequent educational development? Unpublished Ph.D. dissertation, Syracuse University, May 9.

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

Gilliam, W., & Ripple, C. (2004). What can be learned from state-funded preschool initiatives?: A data-based approach to the Head Start devolution debate. In E. Zigler & S.J. Styfco (Eds.), *The Head Start debates* . Baltimore, MD: Paul H. Brookes Publishing, 477-497.

Gilliam, W., & Zigler, E. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation, *Early Childhood Research Quarterly, 15,* 441-473.

Gilliam, W., & Zigler, E. (2004). *State efforts to evaluate the effects of prekindergarten: 1977 to 2003.* Unpublished manuscript: Yale University Child Study Center.

Goodson, B. & Moss, M. (1992). Analysis report: Predicting quality of early childhood classrooms. *Observational Study of Early Childhood* Programs. Boston, MA: Abt Associates, Inc.

Gormley, W., & Gayer, T. (2005) Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program, *Journal of Human Resources* (Summer).

Gormley, W., & Phillips, D. (2005). The effects of universal pre-K in Oklahoma: Research highlights and policy implications, *Policy Studies Journal* (February).

Henry, G., Gordon, C., Mashburn, A. & Ponder, B. (2001). *Pre-K Longitudinal Study: Findings from the 1999-2000 School Year.* Atlanta, Georgia: Georgia State University, Applied Research Center.

Henry, G.T., Henderson, L.W., Ponder, B.D., Gordon, C.S., Mashburn, A.J., & Rickman, D.K. (2003). *Report of the Findings from the Early Childhood Study: 2001-02.* Altanta, GA: Georgia State University, Andrew Young School of Policy Studies.

Loeb, S., Fuller, B., Kagan, S.L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *75*(1), 47-65.

Love, J. (in press). Early Head Start and early development. In K. McCarthy & D. Phillips (Ed), *Handbook of Early Child Development* Oxford, UK: Blackwell Publishers.

Love, J.M., Kisker, E.E., Ross, C.M., Brooks-Gunn, J., Schochet, P., Boller, K., Constantine, J., Vogel, C., Fuligni, A., & Brady-Smith, C. (2002). Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Report prepared for the Administration for Children and Families, U.S. Department of Health and Human Services. Princeton, NJ: Mathematica Policy Research.

McCarton, C.M., Brooks-Gunn, J., Wallace, I.F., & Bauer, C.R. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants: The Infant health and development program. *Journal of the American Medical Association, 277*(2), 126-132.

Magnuson, K., Meyers, M., Ruhm, C., and Waldfogel, J. (2003). *Inequality in Preschool Education and School Readiness.* New York: Columbia University, School of Social Work.

Magnuson, K., Ruhm, C., and Waldfogel, J. (2004). *Does Prekindergarten Improve School Preparation and Performance?* Cambridge, Mass.: National Bureau of Economic Research, April.

Mather, N., & Woodcock, R. (2001). *Examiner's manual, Woodcock-Johnson Achievement Test –III.* Chicago: Riverside Publishing Co.

NICHD Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89*, 1072-1077.

NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71,* 960-980.

NICHD Early Child Care Research Network. (2002). Child care structure, process, outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science, 13,* 199-206.

NICHD Early Child Care Research Network & Duncan, G. Y. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74,* 1454-1475.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C. Kagan, S.L. et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*, 1534-1553.

Phillips, D.A., Voran, M., Kisker, E., Howes, C., & Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? *Child Development, 65,* 472-492.

Pianta, R.C., & Rimm-Kaufmann, S. (in press). The social ecology of the transition to school: Classrooms, families, and children. In K. McCarthy & D. Phillips (Ed), *Handbook of Early Child Development*. Oxford, UK: Blackwell Publishers.

Ramey, C.T., Bryant, D.M., & Suarez, T.M. (1985). Preschool compensatory education and the modifiability of intelligence: A critical review. In D.K. Detterman (Ed.), *Current topics in human intelligence. Vol. 1. Research methodology*. Norwood, N.J.: Ablex Publishing Corp., 247-296.

Ramey, C.T., Bryant, D.M., Wasik, B.H., Sparling, J.J., Fendt, K.H., & LaVange, L.M. (1992). Infant Health and Development Program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics* 89 (3): 454-65.

Ramey, C.T., Campbell, F.A., Burchinal, M., Skinner, M.L., Gardner, D.M., & Ramey, S.L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science* 4 (1): 2-14.

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

Schweinhart, L., Barnes, H., Weikart, D., Barnett, W.S., & Epstein, A. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Vol. 10). Ypsilanti, MI: High/Scope Press.

Stipek, D. & Byler, P. (2003). The early childhood classroom observation measures. Unpublished manuscript. Graduate School of Education, Stanford University.

U.S. Government Accountability Office (2004). *Prekindergarten: Four selected states expanded access by relying on schools and existing providers of early education and care to provide* services. Washington, D.C.: September 9.

Votruba-Drzal, E., Coley, R.L., & Chase-Lansdale, P.L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development, 75*(1), 296-312.

Xiang, Z., & Schweinhart, L. (2002). *Effects Five Years Later: The Michigan School Readiness Program Evaluation Through Age 10.* Paper prepared for the Michigan State Board of Education. Ypsilanti, Michigan: High/Scope Educational Research Foundation, January 23.

Author Note

William T. Gormley, Jr., Public Policy Institute, Georgetown University; Ted Gayer, Public Policy Institute, Georgetown University; Deborah Phillips, Department of Psychology, Georgetown University; Brittany Dawson, Center for Research on Children in the U.S., Georgetown University.

This research was conducted by scholars affiliated with the Center for Research on Children in the U.S. (CROCUS) at Georgetown University. We received helpful comments from Steve Barnett, Michael Foster, Michael Lopez, Ruby Takanishi, and two anonymous reviewers. We would like to thank Naomi Keller, Berkeley Smith, Cynthia Schuster, Preethi Hittalmani, Helen Cymrot, Bonnie Gordic, Alexis Lester, Belen Rodas, Elizabeth Rutzick, Jennifer Scott, Ria Sengupta, Samantha Sydnor, Lindsay Warner, Kristie Carpec, and Branden Kelly for their invaluable research assistance. We greatly appreciate the cooperation we have received from the Tulsa Public Schools and its teachers and principals. Finally, we would like to thank the Foundation for Child Development, the National Institute for Early Education Research (NIEER), and the Pew Charitable Trusts for their generous financial support. We alone are responsible for the contents of this report.

Correspondence concerning this article should be addressed to: William T. Gormley, Jr., Public Policy Institute, Georgetown University, 3520 Prospect St., Washington, D.C. 20007. E-mail: gormleyw@georgetown.edu.

Figure 1: Regression Discontinuity Design with Effective Treatment

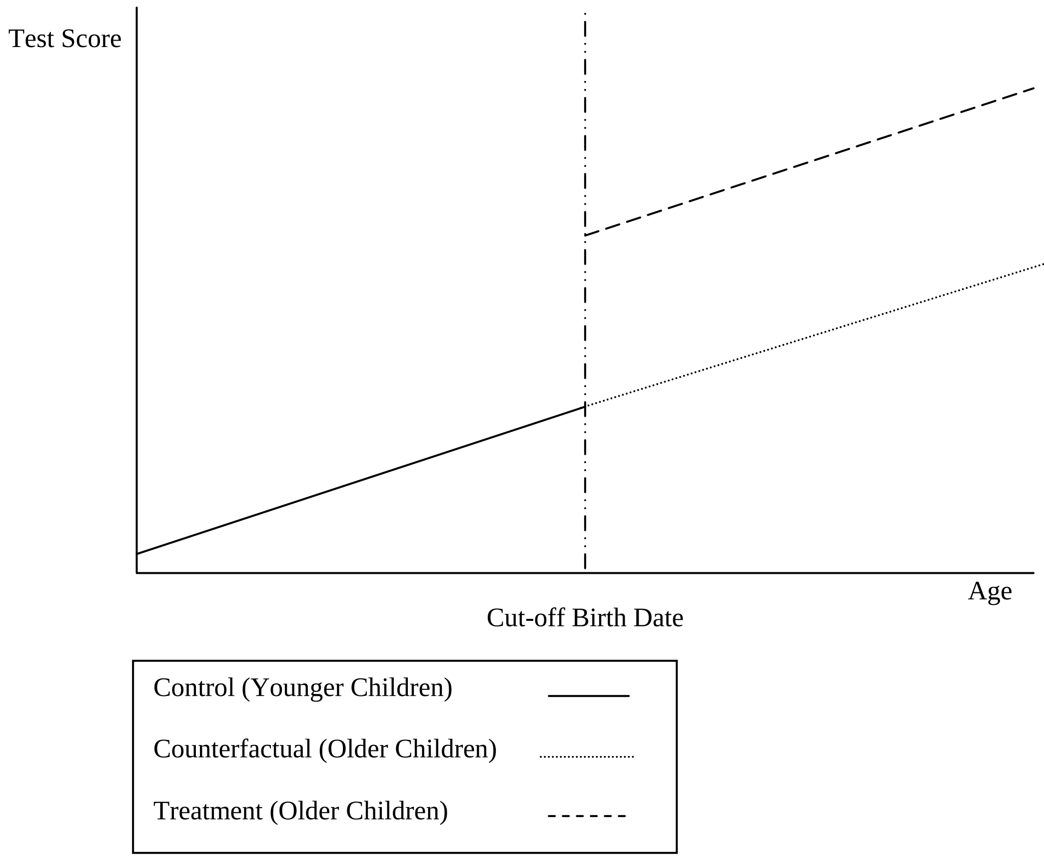

TABLE 1

COMPARISON OF TESTED CHILDREN, UNIVERSE OF CHILDREN

	Pre-Kindergarten			Kindergarten		
	Tested Children	**Universe**	**Difference of Means**	**Tested Children**	**Universe**	**Difference of Means**
Male	50.3% (0.50)	50.9% (0.50)	0.6%	53.0% (0.50)	53.4% (0.50)	0.4%
N	1,567	1,845		3,149	3,727	
White	35.2% (0.48)	36.4% (0.48)	1.2%	40.2% (0.49)	38.3% (0.49)	-1.9%
Black	38.7% (0.49)	35.9% (0.48)	-2.8%[c]	32.6% (0.47)	32.7% (0.47)	0.1%
Hispanic	16.8% (0.37)	17.8% (0.38)	1.0%	15.0% (0.36)	17.3% (0.38)	2.3% [a]
Native American	8.1% (0.27)	8.7% (0.28)	0.6%	10.9% (0.31)	10.3% (0.30)	-0.6%
Asian	1.2% (0.11)	1.2% (0.11)	0.0%	1.3% (0.42)	1.3% (0.50)	0.0%
N	1,567	1,843		3,149	3,727	
Free Lunch	54.6% (0.50)	54.3% (0.50)	-0.3%	54.5% (0.50)	59.4% (0.50)	4.9% [a]
Reduced Price	10.6% (0.31)	10.3% (0.30)	-0.3%	8.7% (0.28)	8.2% (0.28)	-0.5%
Full Price	34.8% (0.48)	35.6% (0.48)	-0.6%	36.8% (0.48)	32.4% (0.46)	-4.4% [a]
N	1,525	1,864		3,087	3,782	

Note: Standard deviations are in parentheses. The tested children are those children who were tested and whose test results were intelligible. Most children were tested in September 2003. The universe of children represents those children who were enrolled in Tulsa Public Schools as of October 2003. Neither the tested children nor the universe of children for the pre-K cohort includes children in Head Start programs, most of which collaborate with Tulsa Public Schools.

[a] $p < .01$, [b] $p < .05$, [c] $p < .10$

Table 2: Comparison of Mean Test Scores and Covariates for K Children Who Were in TPS Pre-K vs. K Children Who Were Not in TPS Pre-K

Variable	|Margin|<=1 year				
	No Pre-K		Pre-K		Difference
Letter-Word Score	7.455		10.439		2.984
	(0.126)	N=1587	(0.145)	N=1349	P=0.0000
Spelling Score	8.713		9.969		1.256
	(0.074)	N=1575	(0.079)	N=1344	P=0.0000
App. Prob. Score	13.374		14.730		1.356
	(0.130)	N=1585	(0.131)	N=1348	P=0.0000
Full Price Lunch	0.395		0.335		-0.060
	(0.012)	N=1538	(0.013)	N=1339	P=.0009
Reduced Price	0.068		0.110		0.042
Lunch	(0.006)	N=1538	(0.009)	N=1339	P=0.0001
Free Lunch	0.536		0.555		0.018
	(0.013)	N=1538	(0.014)	N=1339	P=0.3209
Non-White	0.589		0.629		0.039
	(0.012)	N=1587	(0.013)	N=1349	P=0.0292
White	0.411		0.371		-0.039
	(0.012)	N=1587	(0.013)	N=1349	P=0.0292
Black	0.297		0.391		0.094
	(0.011)	N=1587	(0.013)	N=1349	P=0.0000
Hispanic	0.172		0.113		-0.0593
	(0.009)	N=1587	(0.009)	N=1349	P=0.0000
Native American	0.109		0.110		0.0007
	(0.008)	N=1587	(0.009)	N=1349	P=0.9517
Asian	0.011		0.015		0.004
	(0.003)	N=1587	(0.003)	N=1349	P=0.3195
Female	0.483		0.469		-0.013
	(0.013)	N=1587	(0.014)	N=1349	P=0.4678
Mother No H.S.	0.235		0.167		-0.068
	(0.012)	N=1343	(0.011)	N=1160	P=0.0000
Mother Educ. - High School	0.416		0.446		0.029
	(0.013)	N=1343	(0.015)	N=1160	P=0.1378
Mother Some Coll.	0.189		0.222		0.033
	(0.011)	N=1343	(0.012)	N=1160	P=0.0395
Mother College Deg.	0.159		0.165		.0053
	(0.010)	N=1343	(0.011)	N=1160	P=0.7192

Notes: Standard errors in parentheses.

Table 3: Comparison of Mean Test Scores and Covariates Before and After Cut-Off Birth Date

(Kindergarten Children are Conditional on Having Been in TPS Pre-kindergarten)

Variable	|Margin|<=1 year					|Margin|<=6 months					|Margin|<=3 months					Quadratic Parametric Fit				
	After 9/1 (Control)		Before 9/1 (Treatment)		Diff.	After 9/1 (Control)		Before 9/1 (Treatment)		Diff.	After 9/1 (Control)		Before 9/1 (Treatment)		Diff.	After 9/1 (Control)		Before 9/1 (Treatment)		Diff.
Letter-Word	4.548		10.413		5.865	5.153		9.534		4.381	5.409		8.848		3.439	5.656		8.704		3.048
Score	-.0986	N=1490	(0.145)	N=1355	P=0.0000	(0.146)	N=753	(0.184)	N=659	P=0.0000	(0.208)	N=388	(0.242)	N=310	P=0.0000	(0.282)	N=1490	(0.442)	N=1355	P=0.0000
Spelling	5.657		9.9511		4.294	6.388		9.392		3.003	6.755		8.945		2.190	6.946		8.844		1.879
Score	(0.076)	N=1464	(0.079)	N=1350	P=0.0000	(0.109)	N=739	(0.109)	N=656	P=0.0000	(0.158)	N=380	(0.163)	N=310	P=0.0000	(0.210)	N=1464	(0.239)	N=1350	P=0.0000
App. Prob.	8.649		14.705		6.056	9.636		13.697		4.061	10.230		13.055		2.825	10.632		12.594		1.962
Score	(0.133)	N=1488	(0.131)	N=1354	P=0.0000	(0.192)	N=752	(0.177)	N=659	P=0.0000	(0.274)	N=387	(0.254)	N=310	P=0.0000	(0.376)	N=1488	(0.395)	N=1354	P=0.0003
Full Price	0.351		0.336		-0.015	0.365		0.352		-0.013	0.353		0.324		-0.029	0.341		0.342		0.001
Lunch	(0.012)	N=1460	(0.013)	N=1342	P=0.3944	(0.018)	N=732	(0.019)	N=653	P=0.6277	(0.025)	N=377	(0.027)	N=306	P=0.4228	(0.036)	N=1460	(0.040)	N=1342	P=0.9891
Reduced Price	0.107		0.110		0.003	0.107		0.118		0.011	0.135		0.108		-0.027	0.132		0.107		-0.025
Lunch	(0.008)	N=1460	(0.009)	N=1342	P=0.8190	(0.011)	N=732	(0.013)	N=653	P=0.5036	(0.017)	N=377	(0.018)	N=306	P=0.2783	(0.024)	N=1460	(0.026)	N=1342	P=0.4789
Free Lunch	0.542		0.554		0.013	0.529		0.530		-0.001	0.512		0.569		-0.057	0.527		0.552		0.024
	(0.013)	N=1460	(0.014)	N=1342	P=0.5029	(0.018)	N=732	(0.020)	N=653	P=0.9652	(0.026)	N=377	(0.028)	N=306	P=0.1399	(0.038)	N=1460	(0.042)	N=1342	P=0.6677
Non-White	0.647		0.628		-0.018	0.649		0.625		0.024	0.665		0.648		0.017	0.687		0.657		-0.030
	(0.012)	N=1490	(0.013)	N=1355	P=0.3132	(0.017)	N=753	(0.019)	N=659	P=0.3452	(0.024)	N=388	(0.027)	N=310	P=0.6474	(0.036)	N=1490	(0.041)	N=1355	P=0.5746
White	0.353		0.371		-0.018	0.351		0.375		0.024	0.335		0.352		-0.017	0.313		0.343		0.031
	(0.012)	N=1490	(0.013)	N=1355	P=0.3132	(0.017)	N=753	(0.019)	N=659	P=0.3452	(0.024)	N=388	(0.027)	N=310	P=0.6474	(0.036)	N=1490	(0.041)	N=1355	P=0.5746
Black	0.377		0.393		0.015	0.382		0.393		0.011	0.363		0.419		0.056	0.378		0.434		0.056
	(0.013)	N=1490	(0.013)	N=1355	P=0.3981	(0.018)	N=753	(0.019)	N=659	P=0.6850	(0.024)	N=388	(0.028)	N=310	P=0.1322	(0.037)	N=1490	(0.041)	N=1355	P=0.3077
Hispanic	0.175		0.112		-0.063	0.175		0.115		0.060	0.193		0.116		-0.077	0.208		0.128		-0.081
	(0.010)	N=1490	(0.009)	N=1355	P=0.0000	(0.014)	N=753	(0.012)	N=659	P=0.0015	(0.020)	N=388	(0.018)	N=310	P=0.0056	(0.029)	N=1490	(0.027)	N=1355	P=0.0392
Native	0.083		0.109		0.027	0.081		0.105		0.024	0.093		0.100		0.007	0.087		0.073		-0.014
American	(0.007)	N=1490	(0.008)	N=1355	P=0.0155	(0.009)	N=753	(0.012)	N=659	P=0.1246	(0.015)	N=388	(0.017)	N=310	P=0.7482	(0.021)	N=1490	(0.026)	N=1355	P=0.4165
Asian	0.012		0.015		0.003	0.106		0.012		-0.002	0.015		0.013		0.003	0.013		0.02		0.007
	(0.003)	N=1490	(0.003)	N=1355	P=0.5342	(0.004)	N=753	(0.004)	N=659	P=0.7886	(0.006)	N=388	(0.006)	N=310	P=0.7777	(0.008)	N=1490	(0.010)	N=1355	P=0.6029
Female	0.499		0.469		0.030	0.477		0.463		-0.014	0.479		0.455		-0.025	0.454		0.448		-0.006
	(0.013)	N=1490	(0.014)	N=1355	P=0.1096	(0.018)	N=753	(0.019)	N=659	P=0.6009	(0.025)	N=388	(0.028)	N=310	P=0.5192	(0.038)	N=1490	(0.042)	N=1355	P=0.9156
Mother No	0.236		0.169		-0.067	0.223		0.180		-0.043	0.257		0.165		-0.092	0.302		0.174		-0.128
H.S.	(0.012)	N=1358	(0.011)	N=1164	P=0.0000	(0.016)	N=686	(0.016)	N=571	P=0.0616	(0.023)	N=354	(0.023)	N=267	P=0.0057	(0.033)	N=1358	(0.034)	N=1164	P=0.0072
Mother H.S.	0.397		0.444		0.047	0.404		0.440		0.036	0.384		0.464		0.080	0.371		0.433		0.063
Degree	(0.013)	N=1358	(0.015)	N=1164	P=0.0165	(0.019)	N=686	(0.021)	N=571	P=0.2008	(0.026)	N=354	(0.031)	N=267	P=0.0449	(0.039)	N=1358	(0.045)	N=1164	P=0.2930
Mother Some	0.233		0.223		-0.011	0.235		0.221		-0.014	0.215		0.240		0.025	0.203		0.260		0.057
College	(0.011)	N=1358	(0.012)	N=1164	P=0.5150	(0.016)	N=686	(0.017)	N=571	P=0.5556	(0.022)	N=354	(0.026)	N=267	P=0.4611	(0.033)	N=1358	(0.038)	N=1164	P=0.2563
Mother Coll.	0.133		0.164		0.031	0.138		0.159		0.021	0.144		0.131		-0.013	0.124		0.133		0.009
Degree	(0.009)	N=1358	(0.011)	N=1164	P=0.0296	(0.013)	N=686	(0.015)	N=571	P=0.2995	(0.019)	N=354	(0.021)	N=267	P=0.6435	(0.027)	N=1358	(0.033)	N=1164	P=0.8395

Notes: Standard errors in parentheses.

Table 4: The Effect of TPS Pre-K on Test Scores:
Quadratic Parametric Fit
(Kindergarten Children are Conditional on Having Been in TPS Pre-kindergarten)

Variable	|Margin|<=1 Year			|Margin|<=6 Months			|Margin|<=3 Months		
	Letter-Word	Spelling	Applied Problems	Letter-Word	Spelling	Applied Problems	Letter-Word	Spelling	Applied Problems
Born Before Cut-Off (Treated)	2.999 [a]	1.857 [a]	1.939 [a]	2.231 [a]	0.889 [b]	1.261 [c]	2.651 [a]	0.767	2.253 [b]
	(0.501)	(0.324)	(0.506)	(0.721)	(0.447)	(0.705)	(1.028)	(0.654)	(0.989)
Qualify (days)	0.007 [c]	0.008 [a]	0.016 [a]	0.018	0.031 [a]	0.029 [b]	0.001	0.042 [c]	0.005
	(0.004)	(0.003)	(0.005)	(0.012)	(0.008)	(0.013)	(0.032)	(0.024)	(0.038)
$Qualify^2$	1.9E-06	-1.4E-07	1.4E-05	6.2E-05	1.0E-04 [a]	8.0E-05	-1.3E-04	3.0E-04	-1.0E-04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Qualify * Cut-Off	-0.002	-0.005	-0.012 [c]	0.003	-0.018	-0.016	0.021	-0.022	-0.028
	(0.006)	(0.004)	(0.007)	(0.018)	(0.011)	(0.018)	(0.050)	(0.034)	(0.052)
$Qualify^2$ * Cut-Off	1.1E-05	7.2E-06	-3.4E-06	-1.4E-04	-1.7E-04 [a]	-1.1E-04	-8.3E-05	-4.0E-04	4.5E-04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Reduced Price Lunch	-0.964 [a]	-0.290 [c]	-0.454	-0.738 [b]	-0.111	-0.382	-0.463	-0.116	-0.706
	(0.262)	(0.175)	(0.288)	(0.371)	(0.262)	(0.406)	(0.488)	(0.381)	(0.601)
Free Lunch	-1.283 [a]	-0.885 [a]	-1.379 [a]	-1.227 [a]	-0.792 [a]	-1.152 [a]	-0.876 [b]	-0.570 [b]	-1.040 [b]
	(0.212)	(0.132)	(0.206)	(0.293)	(0.189)	(0.302)	(0.400)	(0.279)	(0.436)
Black	0.037	-0.441 [a]	-2.339 [a]	0.171	-0.599 [a]	-2.548 [a]	-0.029	-0.856 [a]	-2.431 [a]
	(0.211)	(0.128)	(0.204)	(0.301)	(0.185)	(0.279)	(0.405)	(0.295)	(0.422)
Hispanic	-1.700 [a]	-0.482 [a]	-3.656 [a]	-1.388 [a]	-0.576 [b]	-3.833 [a]	-1.243 [a]	-0.623	-4.357 [a]
	(0.240)	(0.180)	(0.332)	(0.373)	(0.274)	(0.507)	(0.506)	(0.386)	(0.682)
Native American	-0.253	-0.105	-0.555 [c]	-0.091	-0.104	-0.527	-0.109	-0.075	-0.568
	(0.290)	(0.193)	(0.324)	(0.409)	(0.288)	(0.452)	(0.580)	(0.406)	(0.630)
Asian	2.976 [b]	1.425 [b]	-2.918 [a]	0.969	1.265 [c]	-4.636 [a]	0.819	1.772 [c]	-5.308 [a]
	(1.498)	(0.696)	(1.070)	(1.201)	(0.717)	(1.747)	(1.239)	(1.060)	(2.094)
Female	0.915 [a]	1.046 [a]	0.761 [a]	1.085 [a]	1.188 [a]	0.637 [a]	1.198 [a]	1.212 [a]	0.753 [b]
	(0.168)	(0.105)	(0.170)	(0.239)	(0.152)	(0.243)	(0.316)	(0.229)	(0.352)
Mother H.S. Deg.	0.591 [a]	0.567 [a]	1.252 [a]	0.742 [a]	0.624 [a]	1.319 [a]	0.849 [b]	0.585 [c]	0.960 [c]
	(0.196)	(0.146)	(0.251)	(0.292)	(0.214)	(0.369)	(0.420)	(0.322)	(0.516)
Mother Some Coll.	1.747 [a]	0.966 [a]	2.199 [a]	1.564 [a]	0.901 [a]	2.308 [a]	1.999 [a]	1.126 [a]	2.224 [a]
	(0.246)	(0.170)	(0.287)	(0.346)	(0.244)	(0.414)	(0.499)	(0.371)	(0.606)
Mother College Deg.	3.274 [a]	1.710 [a]	3.388 [a]	3.407 [a]	1.891 [a]	3.862 [a]	3.588 [a]	2.012 [a]	3.490 [a]
	(0.357)	(0.204)	(0.338)	(0.484)	(0.301)	(0.497)	(0.707)	(0.444)	(0.732)
Number of Obs.	2484	2463	2483	1232	1220	1231	607	600	606

Notes: Robust standard errors in parentheses. Statistical significance at the 1%, 5%, and 10% levels (2-tailed) are denoted by "a," "b," and "c," respectively.

Table 5: The Effect of TPS Pre-K on Test Scores
by Race and Half-Day/Full-Day: Quadratic Parametric Fit
(Kindergarten Children are Conditional on Having Been in TPS Pre-kindergarten)

|Margin|<=1 Year		
Letter-Word	**Spelling**	**Applied Problems**
	Hispanic Children	
4.149 [a]	2.658 [a]	4.969 [a]
(1.108)	(0.790)	(1.432)
N=322	N=310	N=321
	Black Children	
2.911 [a]	1.469 [a]	1.682 [b]
(0.810)	(0.545)	(0.759)
N=969	N=964	N=969
	Native American Children	
3.561 [b]	2.241 [b]	3.081 [c]
(1.738)	(1.151)	(1.788)
N=240	N=239	N=240
	White Children	
3.022 [a]	2.067 [a]	0.852
(0.886)	(0.516)	(0.837)
N=925	N=922	N=925
	Full Price Lunch Children	
2.687 [a]	1.590 [a]	1.543 [c]
(0.927)	(0.532)	(0.887)
N=873	N=872	N=873
	Reduced Price Lunch Children	
5.002 [a]	2.810 [a]	1.403
(1.352)	(0.970)	(1.469)
N=277	N=276	N=277
	Free Lunch Children	
2.791 [a]	1.750 [a]	2.097 [a]
(0.659)	(0.446)	(0.681)
N=1334	N=1315	N=1333
	Half-Day Children	
3.003 [a]	2.265 [a]	2.176 [a]
(0.708)	(0.435)	(0.736)
N=1347	N=1346	N=1346
	Full-Day Children	
2.847 [a]	1.476 [a]	1.655 [a]
(0.712)	(0.477)	(0.695)
N=1134	N=1114	N=1134
	Hispanic, Half-Day Children	
3.700 [a]	3.433 [a]	6.082 [a]
(1.442)	(1.151)	(1.925)
N=207	N=206	N=206
	Hispanic, Full-Day Children	
4.644 [a]	1.693	2.494
(1.505)	(1.068)	(2.416)
N=115	N=104	N=115
	Black, Half-Day Children	
1.189	2.260 [c]	-0.179
(2.039)	(1.289)	(1.771)
N=212	N=212	N=212
	Black, Full-Day Children	
3.090 [a]	1.263 [b]	1.925 [b]
(0.862)	(0.602)	(0.832)
N=756	N=751	N=756
	Native American, Half-Day Children	
3.325 [c]	1.667	3.354
(1.747)	(1.224)	(2.230)
N=156	N=156	N=156
	Native American, Full-Day Children	
3.572	3.222	1.744
(3.804)	(2.324)	(3.140)
N=83	N=82	N=83
	White, Half-Day Children	
3.152 [a]	1.868 [a]	1.062
(1.004)	(0.582)	(0.941)
N=750	N=750	N=750
	White, Full-Day Children	
1.836	3.020 [a]	-0.174
(2.259)	(1.127)	(1.801)
N=174	N=171	N=174

Notes: Robust standard errors in parentheses. Each row represents a different set of regressions, each pertaining to a different race by half-day/full-day designation. Each regression estimation includes the relevant covariates from Table 4. Statistical significance at the 1%, 5%, and 10% levels (two-tailed) are denoted by "a", "b", and "c" respectively.

Footnotes

[1] A child's participation in a particular type of program is based on parental reports.

[2] Personal communication, Dr. Ramona Paul, Assistant Superintendent of Education, Oklahoma, March 14, 2003.

[3]The penetration rates are even higher if one includes participation in Head Start programs. According to the U.S. Government Accountability Office (2004: 12-13), penetration rates (including both publicly-funded pre-K programs and Head Start programs) are: Oklahoma, 74 percent; Georgia, 63 percent; and New York, 39 percent.

[4] Oklahoma City is the second largest school district in the state.

[5] http://www.tulsaschools.org/Profiles/District Summary.pdf, accessed 6/11/04.

[6] We have defined the universe for the Tulsa Pre-K program as all TPS pre-K students, exclusive of Head Start students. An additional 630 students, in Head Start "collaborative" programs, were analyzed separately, and a small number of children in other collaboratives were not analyzed at all.

[7] Barbara Wendling, an independent consultant, based in Dallas, Texas, conducted the training sessions. Wendling is a nationally recognized expert on the Woodcock-Johnson Achievement Test. She is co-author, with Nancy Mather and Richard Woodcock, of *Essentials of WJ III Tests of Achievement Assessment.*

[8] Unfortunately, the relationship between the birthday qualification date and assignment in the treatment group is not perfectly discontinuous. There are four children in kindergarten who should be in Tulsa pre-K based on their birthday. Similarly, there are 24 children in Tulsa pre-K who should be in kindergarten based on their birthday. We conduct our analysis as if the birthday qualification was perfectly enforced, and we drop these 28 observations from our sample. Alternatively, one could keep these aberrational observations and use an instrumental variables approach in which the cut-off dummy variable serves as an instrument for whether the child has just completed Tulsa pre-K. Given the small number of aberrational observations, the results of the two approaches are virtually identical.

[9] Additionally, we computed the percent increase in test scores from the treatment effect by running a regression in which the dependent variable is logged. In order to do this, we first dropped the observations in which test scores equal zero. We find a 50.52% increase in the letter-word score, a 25.20% increase in the spelling score, and a 19.71% increase in the applied problems score.

[10] In summarizing the results of a meta-analysis by Gilliam and Zigler, we have focused on overall development and achievement tests, as they appear in Table 5, Column K.