

OCTOBER 15, 2014 | NUMBER 760

The Evidence on Universal Preschool Are Benefits Worth the Cost?

BY DAVID J. ARMOR

EXECUTIVE SUMMARY

Calls for universal preschool programs have become commonplace, reinforced by President Obama's call for "high-quality preschool for all" in 2013. Any program that could cost state and federal taxpayers \$50 billion per year warrants a closer look at the evidence on its effectiveness.

This report reviews the major evaluations of preschool programs, including both traditional programs such as Head Start and those designated as "high quality." These evaluations do not paint a generally positive picture. The most methodologically rigorous evaluations find that the academic benefits of preschool programs are quite modest, and these gains fade after children enter elementary school. This is the case for Head Start, Early Head Start, and also for the "high-quality" Tennessee preschool program. Meanwhile, most contemporary "high-quality" preschool programs have been evaluated using a flawed, non-experimental methodology called Regression Discontinuity Design (RDD). Existing RDD studies fail to

account for children who drop out of treatment groups, thereby biasing outcomes upwards. Further, by their nature RDD evaluations cannot assess the fadeout problem because all children in the study, both treatment and control, have taken preschool. These problems affect the evaluation of the Tulsa, Oklahoma, program, perhaps the most frequently cited contemporary "high-quality" program.

Two "high-quality" programs have been evaluated using a rigorous experimental design, and have been shown to have significant academic and social benefits, including long-term benefits. These are the Abecedarian and Perry Preschool programs. However, using these two studies as the basis for policy is problematic for several reasons: the groups studied were very small, they came from single communities several decades ago, and both programs were far more intensive than the programs being contemplated today.

Before policymakers consider huge expenditures to expand preschool, especially by making it universal, much more research is needed to demonstrate true effectiveness.

“The most rigorous studies of contemporary preschool programs show no lasting gains for preschool students after they enter regular grades.”

INTRODUCTION

President Obama’s 2013 proposal of “high-quality preschool for all,” with \$75 billion in federal startup money, inspired similar calls for universal preschool by state and local leaders throughout the country. New York City mayor Bill de Blasio ran on a universal pre-K platform, funded by “taxing the rich,” while state governor Andrew Cuomo has also endorsed universal pre-K funded from existing tax monies. Preschool has become a major issue in the gubernatorial races in Texas and Ohio, with both Democratic candidates endorsing universal pre-K. The California legislature is considering a proposal spearheaded by Senate president Darrell Steinberg to expand its coverage of “transitional kindergarten” to all four-year-olds, thereby becoming a universal pre-K program.

Expanding public schooling to cover all four-year-olds would require a significant increase in state and federal spending. Even if Congress authorized the full Obama proposal, which covers a 10-year period, total expenditures for the states would be far higher. With U.S. expenditures for public schooling from all sources exceeding \$12,000 per student, and with approximately 4 million students enrolled in public kindergartens, states could be spending nearly \$50 billion per year to fund universal preschool, assuming that spending levels for preschool are similar to those for higher grades.

With such large proposed expenditures, the benefits of pre-K should be clear and definitive. Yet existing research on preschool programs does not paint a uniformly positive picture, despite what supporters claim. Indeed, the most rigorous studies of contemporary preschool programs, particularly the federal Head Start program and a Tennessee universal program, show no lasting gains for preschool students after they enter regular grades. According to these studies, by the time children reach the early elementary grades, the average preschool student has learned no more than children who were not in preschool.

How are these results reconciled with proclamations, including some from the White

House, about the dramatic benefits of preschool? Are universal preschool advocates simply ignoring scientific evidence? There is a body of research that finds educational benefits from preschool programs, but these studies suffer from one of two problems: Either the preschool programs are not comparable with the programs being proposed today, or they use non-experimental designs that suffer from serious limitations, including an inability to track preschool effects into the early grades.

Supporters of preschool programs have been very selective in their research citations. They extol programs with large pre-K effects as “high quality,” especially some historical programs and several recent pre-K programs in Tulsa, Oklahoma, and Boston, Massachusetts. Meanwhile they ignore the Head Start findings or imply that it is not a high-quality program.¹

The goal of this study is to offer a more balanced review of the major studies in this field, making it clear what we do and do not know about preschool programs, including both their short- and long-term effects on learning and other behaviors.

HISTORICAL PROGRAMS ARE NOT COMPARABLE TO CONTEMPORARY STUDIES

There are three historical “high-quality” preschool programs that have received considerable attention. These are the Abecedarian project in North Carolina, the Perry Preschool program in Michigan, and the Chicago Child-Parent preschool program.² These historical programs differ markedly from contemporary preschools in ways that make them uncomparable to current and proposed “high-quality” preschool programs. In addition, one of them has some serious methodological limitations.

Abecedarian Program

This program and study was conducted at the University of North Carolina in Chapel Hill during the early 1970s. It was a relatively well-designed randomized experiment involv-

ing 111 low-income infants, nearly all black, 57 of whom were in the treatment group. The program involved interventions from infancy (average of four months) to the start of kindergarten. A Child Development Center provided intensive care and education for up to 40 hours a week for 50 weeks. Follow-up studies, both short- and long-term, found significant gains for the treatment group as compared to the control group on various cognitive and IQ tests up to age 21. However, this very intensive and much longer intervention is not comparable to any contemporary statewide or city preschool program, most of which are designed for a single year of pre-K. It is also worth noting that randomization was done prior to asking mothers' permission to participate, and a higher refusal rate by control group mothers might have produced some bias in results.

Perry Preschool

The Perry Preschool program and research took place in Ypsilanti, Michigan, in the mid-1960s, and involved 123 predominantly black children who attended the Perry Elementary School. It was also a randomized experiment, with 59 students assigned to the preschool program, although the research design was compromised somewhat by switching two working mothers to the control group and by some program dropouts. Many years later this program was the subject of a rigorous cost-benefit study conducted by a group of economists led by James Heckman, who undertook major analyses to compensate for the design flaws. They found that the costs of the Perry Preschool program—which would be about \$20,000 per child in current dollars—were more than paid back by higher employment rates, lower rates of crime, and other economic outcomes favoring the preschool group.³ The Heckman findings generated a great deal of publicity, and they were cited extensively by the White House in support of the Obama proposal.

While the Perry Preschool program is more like a contemporary preschool than the Abecedarian project, the Perry program consisted of two years of preschool coupled with

weekly visits with the parent. Because of the home visits, child-teacher ratios were very small, just five or six children per teacher, far lower than contemporary pre-K programs. Given the critical role parents play in child development, home visits were an important feature of Perry Preschool. The very low child-teacher ratios and the home visits make the Perry Preschool experience uncomparable to the type of preschool programs endorsed by President Obama and other political leaders. Therefore, the cost-benefit results for the Perry program cannot be extrapolated to most contemporary preschool programs, even those that are described as “high quality.”

Chicago Child-Parent Centers

The Chicago Child Parent Center program, which started in 1967, also differs in several ways from standard preschool programs. Like Perry Preschool, there is mandatory parent involvement during the preschool program. Unlike Perry and other preschool programs, it also has follow-on components that continue into the higher grades. Also unlike Perry, it has never been evaluated by a randomized experimental design. Instead, control group children have been selected from other low-income schools, and statistical adjustments have been used to create equivalence. However, self-selection bias is hard to eliminate in studies of this type, particularly when a program runs over multiple years. There is a very high likelihood that treatment outcomes are biased upward by attrition or dropouts who do not complete all years of the program.

CONTEMPORARY EXPERIMENTAL STUDIES

Among contemporary evaluations of regular pre-K programs initiated since the year 2000, two stand out because both used randomized experimental designs. These experimental designs are the “gold standard” for evaluation studies in education. Neither study shows long-term effects of preschool. One is the national Head Start Impact Study, and the

“Neither the national Head Start Impact Study nor the statewide Tennessee program shows long-term effects of preschool.”

“The positive effects did not last beyond the kindergarten year, and there were also no effects on reading and math skills at the end of first or third grades.”

other is an evaluation of the statewide Tennessee program. These experimental studies show modest short-term effects of preschool, measured at the end of the preschool period, but the effects fade after children enter the regular elementary grades. The Tennessee study is of particular interest because it would qualify as a “high-quality” preschool program in the same sense as the Tulsa and Boston programs. A third early-childhood program evaluated with an experimental design is Early Head Start, which is modeled after the Abecedarian project. As such, it is a much more extensive (and expensive) intervention than regular pre-K programs, but it will be discussed briefly for the sake of thoroughness.

Head Start

Perhaps the most important of these contemporary studies is the Head Start Impact Study (HSIS), not only because it is national in scope, but also because it follows students through third grade.⁴ Head Start is also the longest-standing preschool program in the country, having begun in 1965 as part of President Lyndon Johnson’s War on Poverty. In 2012 it served nearly one million students for an annual cost of about \$8 billion.

The Impact Study started with a cohort of over 4,500 children who were 3 or 4 years old in the Fall of 2002.⁵ These students applied for admission to a random sample of more than 300 Head Start centers in 23 states. About 2,600 children were randomly assigned to the Head Start condition, while about 1,800 students were assigned to a control group. Some control group students found other types of preschools, and some students assigned to the Head Start condition did not attend. Special techniques were used to assess how much these “misclassifications” affected the overall results.

The HSIS found statistically significant effects of Head Start for both reading and math skills during the preschool years. However, the effects were modest. A school year constitutes about 10 months of learning, and Head Start students gained about 2 months more than the control group. There were also some significant

effects for social behaviors, particularly parents showing more awareness of health care issues.

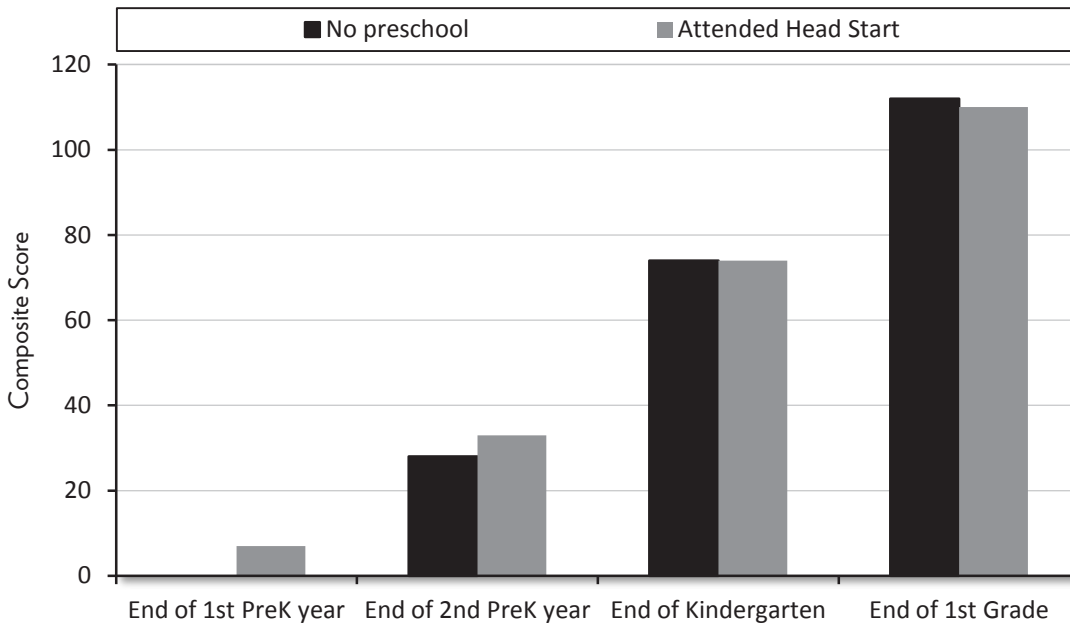
Unfortunately, these positive effects did not last beyond the kindergarten year, and there were also no effects on reading and math skills at the end of first or third grades. Likewise, there were no major effects on problem behaviors lasting into kindergarten, first grade, or third grade.

The HSIS has been criticized because some Head Start students did not finish the Head Start year, and some control group students sought out and entered some other kind of preschool.⁶ The original study defended this as a realistic condition, and in any event it would be unethical to prevent the parents of control group children from seeking out other pre-K programs. Nonetheless, the effect of Head Start might be reduced by these “crossovers,” since some control group children also had preschool. Various statistical analyses were undertaken in the original study to deal with this issue, and a special analysis was undertaken by Peter M. Bernardy.⁷ The results of the Bernardy analysis are shown in Figure 1. In this re-analysis of the HSIS data, none of the control group had any formal preschool, and all of the Head Start students attended the Head Start program for two years. This approach intends to assess the “full” effect of Head Start as compared to children with no preschool experience.

In this special analysis, the Head Start effect at the end of the first pre-K year is 7 points and is statistically significant, representing a 3.5-month advantage for Head Start students over the control (who gained 0 in this analysis). That advantage shrinks to 5 points in the second year of Head Start (about a 2-month advantage) and is still statistically significant. But by the end of kindergarten both Head Start and control students score the same, and by the end of first grade the control students are scoring 2 points higher than Head Start students, although this is not a statistically significant difference.

Note the very large academic skill gains made each year by both groups of students during kindergarten and first grade years. The

Figure 1
Academic Skill Gains for Head Start vs. Control



Source: Peter M. Bernardy, "Head Start: Assessing Common Explanations for the Apparent Disappearance of Initial Positive Effects" (PhD diss., George Mason University, 2012).

7-point gain for three-year-old Head Start students is dwarfed by the gains for both groups during the regular school years, which are about 40 points each year. One of the reasons it is very difficult to realize lasting effects of a preschool program is that children in early elementary school are in their major development years, and they are learning four times as much material during a regular school year as they do in the preschool years, when children are just starting to learn more words, manipulate numbers, and so forth.

Tennessee Pre-K

The Voluntary Pre-K for Tennessee Initiative started in 2005, and by 2007 it had grown to more than 18,000 participants.⁸ The program gives priority to high-risk students, as determined by poverty status (first priority), children with disabilities, and children with limited English proficiency. Like most other state programs, such as those in Tulsa, Oklahoma, and Boston, Massachusetts, preschool teachers must be licensed teachers with pre-K certification, and most classes have teaching

assistants. It is a full-day program, requiring a minimum of 5.5 hours of instruction for five days a week. As such, this program merits the "high quality" label applied to the Tulsa and Boston studies.

Unlike the case in the Tulsa and Boston preschool studies, the evaluation of the Tennessee state preschool program utilized an experimental design whereby 3,000 children were randomly assigned to treatment and control conditions in the 2009–10 and 2010–11 school years. A subgroup of 1,100 children was randomly selected for an "intensive sub-study" involving additional testing and data collection. This study had one major design flaw, in that parental permission for data collection was obtained after the random assignment was made, and the control group had lower participation rates than the treatment group, rendering this no longer a randomized controlled trial but instead only a quasi-experiment. Propensity analysis was used to generate equivalent treatment and control groups.

During the initial evaluation, significant preschool effects were found for academic

“Children in early elementary school are learning four times as much material during a regular school year as they do in the preschool years.”

“The Tennessee randomized (quasi-) experiment had results very similar to those of the national Head Start evaluation.”

achievement, and the magnitudes were modest and similar to effects found of the Head Start Impact Study. For example, the composite academic scale showed an increase of about 2.5 months of learning compared to the control group. At the kindergarten and first grade follow-ups, however, the impacts had diminished. In the words of the investigators,

the effects of [Tennessee Voluntary Pre-K] on the . . . achievement measures observed at the end of the pre-k year had greatly diminished by the end of the kindergarten year and the differences between participants and non-participants were no longer statistically significant. . . . Similarly, at the end of first grade, there were no statistically significant differences between TN-VPK participants and nonparticipants on the [Woodcock-Johnson] measures with one exception . . . that favored the nonparticipant group.⁹

Similarly, at the end of first grade there were no significant differences on a series of behavioral outcome as rated by teachers, including social skills, work-related skills, peer relations, and behavioral problems. In the intensely studied subgroup there was, however, a somewhat higher rate of retention in kindergarten for the control group, 6 percent versus 4 percent (it was 8 percent versus 4 percent for the full 2009–10 cohort).

In contrast to the non-experimental designs used for the Tulsa and Boston evaluations, the Tennessee randomized (quasi-) experiment had results very similar to those of the national Head Start evaluation. There were significant benefits during the preschool year, but the effects faded when students transitioned to the regular school grades, during which time nonpreschool children caught up.

Early Head Start

Early Head Start is a program designed for pregnant women and infants, providing in-home and center-based education and health

services for both the child and the parent(s), extending from the prenatal period until age three.¹⁰ The experimental evaluation of Early Head Start involved 3,000 children and their families who entered the program between 1996 and 1998 and were randomly assigned to a control group and an Early Head Start treatment group. They were followed up at age three, age five, and then at grade five. The program is significantly more costly than regular Head Start; one study estimated annual costs in 2009 to be more than \$11,000 per child/family per year, so that the full three years would cost well over \$30,000 per child.¹¹

The outcomes at fifth grade were not very impressive, particularly given the cost of this program. Out of a total of more than 56 academic, socio-emotional, parenting, behavioral, and mental health outcomes, only two were statistically significant and favored Early Head Start: a socio-emotional success index and an anxiety/depression rating. The sizes of the impacts were quite small, being just one-tenth of a standard deviation. There were no overall effects on cognitive performance (a major objective of all preschool education programs), good parenting behaviors, or reduced negative behaviors such as aggression and rule-breaking.

There were significant positive impacts on several outcomes for the African American subgroup, mainly in the socio-emotional domain and improved parenting behaviors (e.g., involvement in school and less alcohol use). There were no significant impacts on African American cognitive outcomes, which is disappointing to those who hoped that earlier intervention might help close the achievement gap. Perhaps most disappointing was that the highest-risk subgroup (including all races)—the children of single mothers, on welfare, unemployed, and lacking a high school diploma—had no positive outcomes and numerous significant negative impacts, especially in the cognitive domain.

In short, there is little evidence that, even if policymakers are willing to spend the money, Early Head Start is the model that will pro-

duce substantial and consistent improvement in social and academic behavior for disadvantaged children.

PROBLEMS WITH CONTEMPORARY NON-EXPERIMENTAL STUDIES

Several contemporary preschool programs have been evaluated over the past 10 years and have been cited in many press accounts as evidence of the success of “high-quality” universal preschool programs. These include state-sponsored preschool programs involving multiple schools in New Jersey and Georgia, and single-city programs in Tulsa, Oklahoma, and Boston, Massachusetts.¹² An evaluation of statewide universal programs in five states also used the same methodology.

None of these studies used experimental designs, meaning randomized assignment of children to preschool and nonpreschool conditions. Instead, all of these studies used “regression discontinuity designs,” or RDD. The treatment group consisted of children just starting kindergarten who completed preschool the year before, while the control group consisted of children just starting preschool; testing was done at the beginning of the school year for both groups. The design required a strict birth month cutoff (usually September 1) so that children who were four years old as of that date were eligible for preschool, while younger children had to wait until the following year. RDD also assumes that these two groups are identical except for their age and the fact that one has completed preschool while the other has not. The studies used regression analysis to control for the age difference, since age is strongly correlated with test scores.

The Institute for Education Sciences in the U.S. Department of Education publishes standards for RDD studies that must be met in order to make a valid causal inference about treatment effects.¹³ At least one of these standards, and perhaps two, have been violated by the published RDD studies of preschool. First,

if there is any attrition from either group, it must be recorded and reported. Adjustments can be made if effects are small, otherwise valid inferences cannot be drawn. Attrition was not reported for either the Tulsa or the New Jersey RDD studies; it was reported for the Boston studies, but the level of bias was not evaluated as required by the standards.

The attrition problem is particularly important for RDD studies because attrition occurs only for the treatment group. Children who start preschool but who do not complete the program are no longer in the treatment group and are therefore not tested at the beginning of kindergarten. Since the control group children are just starting preschool, there can be no attrition for the control group. Program dropouts, whether due to mobility or difficulty with the preschool program content, are likely to have lower test scores than those who remain in the program, and therefore scores of the treated group can be biased upward by an unknown amount. Moreover, there are no pre-treatment scores for the outcome measures, which could be used to make adjustments if attrition is not too large.

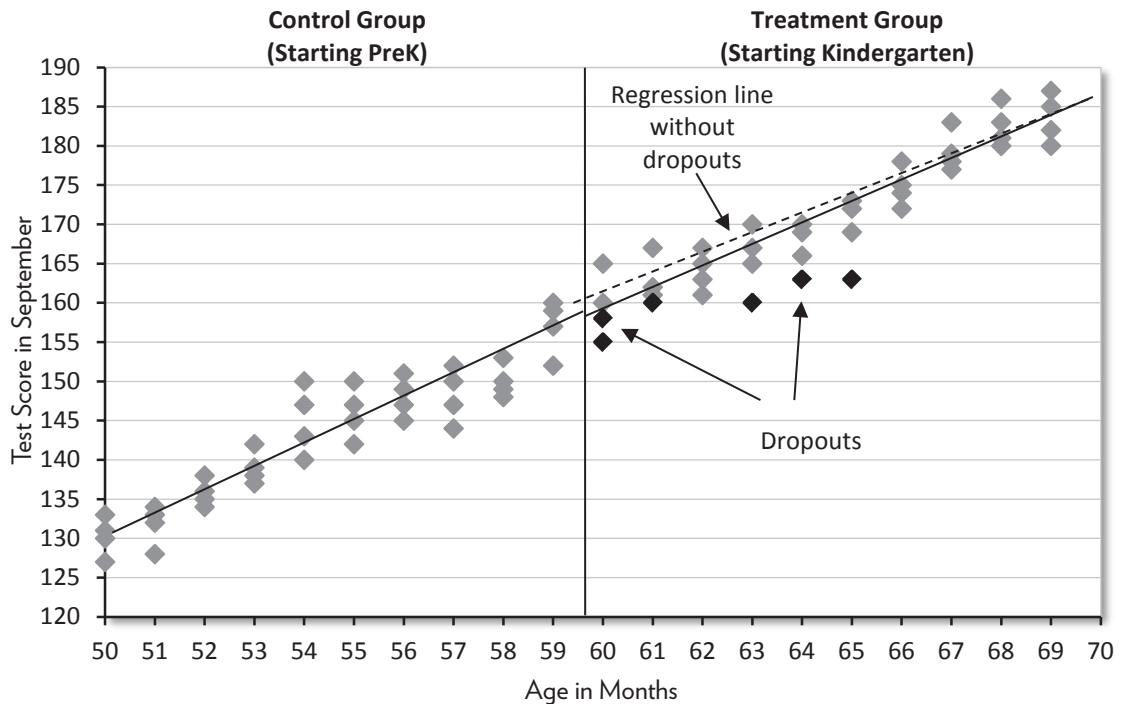
This problem is illustrated using hypothetical data in Figure 2. The solid regression lines for the treatment and control groups show a slight negative effect for treatment (about minus 1 point on the test score). The dotted line shows the treatment regression line after removing test scores for five dropouts (about 12 percent), which now indicates a statistically significant positive effect of about three points.

The second requirement for a reliable regression discontinuity study is that the “forcing” variable used to assign children to treatment and control groups, in this case the cut-off age, must not be confounded with any other characteristic that might cause different behaviors in the treatment and control group. It is likely that the age variable affects many other variables, such as the friends a child plays with, the amount of time spent in out-of-home care, and activities with parents. As a result, age is a problematic choice of forcing variable for the study of pre-school effects.

“Several contemporary preschool programs have been evaluated over the past 10 years. None of them used experimental designs, meaning randomized assignment of children to preschool and nonpreschool conditions.”

“None of the studies provide a rigorous assessment of either the short- or long-term effects of preschool.”

Figure 2
Hypothetical Example of RDD Method



Source: Author's calculations using hypothetical data.

A third problem, which is not mentioned in published standards, is the timing of skill testing. There is no problem if all children are tested on the very first day of schooling or within a few days of the start of school. This may not be realistic, however, since testing can take up to two months or more to complete. The problem arises if the control group testing is completed in a shorter interval than the treatment group, which is possible given the different types of classroom settings. Since cognitive growth rates are much higher in kindergarten than preschool (see Figure 1), this could result in an upward bias of treatment effects. There might be ways to ameliorate this problem, but existing studies do not report on the timing or duration of testing for the two groups.

A fourth problem, and perhaps the most important for preschool studies, is that an RDD study can only assess the effect of preschool for the year of treatment; it cannot compare the treatment and control group for the fadeout problem in later grades. Unlike an experimen-

tal design, both the treatment and the control group will have had preschool, and thus there is no “identical” group that has not had preschool. The long-term effect of preschool is the critical policy issue, because there is little educational value for preschool effects that occur during the preschool year but that do not extend into the later grades, as when a 1st- or 2nd-grade student who has had preschool is compared to a student who has not.¹⁴

Because of these problems, none of the studies discussed below provide a rigorous assessment of either the short- or long-term effects of preschool. They are reviewed here because many commentators (including the studies' authors) consider them to demonstrate the efficacy of high-quality programs.

The Abbot Preschool Program:

The New Jersey Abbot Preschool program is offered in 31 predominantly low-income school districts and serves more than 40,000 three- and four-year-old children, which con-

stitutes approximately 80 percent of the eligible children in those districts. The program was implemented in 1999 as a result of a court order. A major RDD evaluation was carried out for approximately 1,500 children, half of whom attended kindergarten in 2005–06 (the treatment group) and half of whom attended preschool (the control group). No information was provided on the number of children who had started preschool the year before but had not completed the program, nor was information available about when the testing was completed for the two groups. This information may not be so important for the Abbot program, however, because the effects of preschool were only slightly larger than those found in the national Head Start program (about 3 months of learning for verbal skills and 3.5 months for math).¹⁵

The Tulsa Program

Oklahoma established a universal, voluntary, pre-K program for four-year-olds in 1998, and at the time of the Tulsa study about 63 percent of its four-year-old population was enrolled in the state program; another 10 percent or so was enrolled in Head Start. This gives Oklahoma the highest rate of pre-K enrollment in the nation; Georgia is a close second with a total of 63 percent total enrolled. The Tulsa study reviewed here was an RDD evaluation for approximately 3,000 students enrolled in pre-K or kindergarten in September of 2003.¹⁶

No detailed information is provided about how long the testing took, except for a statement that “testing took place, for the most part, during the first week of school.” No information is available, either, about the number of students who started pre-K but dropped out before completing the year and thus are not counted as part of the treatment group. Some tabulations suggest that this might be a major issue. For example, the pre-K control group was 18 percent Hispanic, and 26 percent of mothers were high school dropouts, while the kindergarten treatment group was 11 percent Hispanic, and only 16 percent of moth-

ers were high school dropouts. Both of these characteristics are strongly correlated with test scores. These differences could have been created by children from these backgrounds being more likely to drop out of the pre-K program. While these characteristics were statistically taken into account in the regression analysis, the differences suggest the treated and control groups were not “identical” and could differ on unmeasured (and therefore uncontrolled) characteristics, especially motivation and initial cognitive skills.

The impact of the Tulsa preschool program on verbal skills for this cohort of children was 8 months, or almost a full year of schooling; a later evaluation reported an effect of one full year. This is truly a staggering effect; it is unprecedented for any type of school program. It is four times the typical Head Start effect, and greater than any other “high-quality” program, including the Abecedarian and Perry Preschool programs, both of which were far more intensive. It is likely that this effect is an artifact of the RDD method, especially a high dropout rate from the treatment group. This conclusion is reinforced by another study of Oklahoma children that did not produce such extraordinary results (see the five-state study below).

The Boston Program

The Boston preschool program differs to some extent because it was not a statewide program; rather, it was developed by the Boston Schools working with researchers at the Harvard Graduate School of Education. It was initiated in 2005 but was not fully implemented and operational for several years. The evaluation took place for students who attended preschool in 2008–09; testing commenced the following year (September 2009) when the treatment group (969 children) was starting kindergarten and the control group (1049 children) was starting preschool. At that time about one-third of eligible four-year-olds attended preschool.

The study reported a sizable attrition (dropout) rate for the treatment group of about 22 percent and stated that the dropouts differed

“It is likely that the impact of the Tulsa preschool program is an artifact of the Regression Discontinuity Design method, especially a high dropout rate from the treatment group.”

“The most important information needed, such as pre-treatment cognitive skills and motivation, is not available in a Regression Discontinuity Design. This can introduce a significant bias in the results.”

significantly from the completion group. Although adjustments were made using available administrative records, the most important information needed for such adjustments, such as pre-treatment cognitive skills and motivation, is not available in a RDD design. This can introduce a significant bias in the results.

Equally important, although children began school in mid-September, testing did not begin until the end of September, and only one-third of the students had been tested by the end of October. It took until the end of November—more than half way through the semester—before most of the testing was completed. This can introduce another source of bias if the treatment group finished testing later than the control group.

The size of the preschool effects on verbal skills were smaller than the Tulsa studies but were still very large, about six months of growth or three times larger than found for the national Head Start evaluation. Given that these results equaled the effects found for the Abecedarian Project after five years of preschool, and given the reported dropout rate, it is likely that the treatment and control groups were not equivalent.

Study of Five States

A special RDD study was undertaken for universal preschool programs in five states: Michigan, New Jersey, Oklahoma, South Carolina, and West Virginia.¹⁷ Random samples were drawn for preschool classrooms, and random samples of students were drawn from the sampled classes. Kindergarten classes were then selected from the same school as the sampled preschool classroom. Children were tested in the early fall of the 2004–05 school year. Sample sizes ranged from 720 to 870 for all states except New Jersey, where the sample was about 2,000. Treatment and control samples were each about half of the total samples.

In the study no information was provided about attrition or dropout from the treatment groups, and there was no detailed information about the time of testing; we know only that it was done in “early fall.” Attrition and time of

testing might be less important for this study, however, because the effects were not large for most states. Significant effects for verbal skills were only found for New Jersey and Oklahoma, and the magnitudes were only modestly higher than Head Start results (3.5 and 3 months, respectively). The effect for the Oklahoma sample is much smaller than that found in the Tulsa study. The Oklahoma sample did not show a significant effect for math, but the Michigan sample produced a very strong effect for math, nearly 5 months. This is puzzling, because Michigan also showed a small negative effect on reading. No significant effects were found for verbal or math skills for South Carolina and West Virginia.

The Georgia Study

The Georgia study is the most recent RDD study to be published.¹⁸ Data for the evaluation were collected in the fall of 2012 for about 1,200 children, 611 of whom had completed pre-K the previous school year (treatment group) and the remaining children (control group) who were just entering pre-K. The children were recruited from a random sample of 90 pre-K classrooms from the nearly 4,000 pre-K classrooms operating in Georgia as of August 2011.

No systematic information is provided for treatment-group attrition, or children who started preschool but dropped out before finishing. The treatment and control groups were similar on most demographic characteristics except race and English language proficiency (based on the pre LAS test). The control group had a somewhat higher proportion of black students (just 4 percentage points), but a much lower rate of English-language proficiency.

Nearly three-fourths of the treatment group were fully fluent in English compared to only 46 percent of the control group, and only 8 percent of the treatment group were limited or non-English speakers compared to 26 percent of the control group. This result renders the groups non-equivalent on a critical skill, and it is evidence that nonfluent English speakers were less likely to finish the pre-K year. Although the study controlled for Eng-

lish fluency in the regression analysis, the non-equivalence on this characteristic raises the possibility of non-equivalence on other characteristics including cognitive skills in general.

Another concern is the timing of assessment. Although Georgia public schools start in early August, the report states that assessment took place between September 21 and December 20, near the end of the first semester. If assessment took longer for the treatment group in kindergarten, this could introduce another factor that would lead to higher test scores for the treatment group.

The effects of the Georgia program resemble those of the Tulsa study, with an effect on verbal skills of a full school year, and an effect of a half-year on math skills. As with the Tulsa study, these extraordinarily large effects likely reflect an upward bias due to non-equivalent treatment and control groups arising from the RDD technique.

A COMMENT ON META-ANALYSES OF PRESCHOOL PROGRAMS

In addition to the specific studies reviewed here, there are several meta analyses of preschool evaluations that combine results from many individual programs and derive overall estimates of preschool effects. Generally these studies place less emphasis on study designs and more on arriving at an average effect size attributable to preschools programs.

One of the most comprehensive of these reviews examined 84 individual studies including all of the studies listed above.¹⁹ It arrived at an overall estimate of about .2 standard deviations during the program year after adjusting for sample sizes. This is about 2 months of a normal school year during the early elementary years. This effect size is very similar to the average Head Start effect as estimated in the HSIS. The review also found that nearly all follow-up studies showed that the effect faded out after children started regular school. In other words, this meta-analysis produced average effects that were no larger than Head Start and tended to fade out in grade school.

THE QUALITY ISSUE: WHAT CONSTITUTES HIGH-QUALITY PRESCHOOL?

Many discussions of preschool distinguish “high-quality” preschools from others that are presumably not high quality. The classic Abecedarian and Perry Preschool programs, as well as the contemporary Tulsa and Boston preschool programs, have been described as high quality. According to the office that oversees the Tennessee preschool program, it is a “national leader in pre-K quality, achieving 9 out of 10 quality standard benchmarks of the National Institute for Early Education Research (NIEER).”²⁰ It is generally understood that, because it is a national program with thousands of pre-K programs, Head Start classrooms vary in quality.

What is high quality? At least one educator has characterized quality as one of the most confusing issues when evaluating pre-K programs.²¹ Although there are many potential dimensions to the quality debate, the two most important factors cited by many researchers are teacher education and pre-K classroom quality. The state programs that have incorporated pre-K into regular public elementary schools generally have teachers with BA degrees, and many with pre-K certificates, because that is required for public school teachers. In contrast, the national Head Start program does not require BA degrees, and in the HSIS only about 30 percent of Head Start teachers had BA degrees while another 30 percent had AA degrees.

The NIEER describes a preschool classroom quality rating system called ECERS-R, which rates preschool classrooms on seven different dimensions of quality including the classroom space and furnishings, instructional activities, teacher-child interactions, materials, and relations with parents. The rating scale ranges from 1 (poor) to 7 (excellent), and a score of 5 is considered “good.” The rating system has been used for many years to rate the quality of preschool programs, including Head Start programs.

“One of the most comprehensive reviews examined 84 individual studies. This meta-analysis produced average effects that were no larger than Head Start and tended to fade out in grade school.”

“It may be that children who did not have preschool simply caught up with those who did.”

Classroom ratings using this instrument are available for Head Start, the New Jersey Abbot Program, and Boston. In the 2007 New Jersey Study, the average classroom quality rating was 4.8, and 44 percent of classrooms were rated “good” or better. In Boston, 48 percent of classrooms were rated “good” or better in 2007, the year before the formal pre-K evaluation started.

For the Head Start programs in the national Impact Study, the average classroom quality ratings were actually higher than for the New Jersey and Boston programs. The average quality was 5.2, and 70 percent of the classrooms were rated “good” or better. In other words, Head Start scored higher in quality than the reportedly better performing Abbott and Boston programs.

The more important question is whether teacher education and assessed classroom quality make a significant difference in a preschooler’s learning. Most middle class parents do a pretty good job in preparing their children for school without a formal curriculum; there are lots of materials available to help. Moreover, for 3- and 4-year-olds there is no clear reason why a BA degree is necessary for promoting growth in basic skills, as might be the case for older children.

This commonsense reasoning is born out by formal research. For example, a special review of several national and state databases on preschool programs did not find a significant relationship between teachers having a BA degree and various learning outcomes for preschoolers.²² Equally important, a new study using national long-term data on infants and children found no relationship between quality of preschool classrooms on learning or social outcomes once a child’s demographic and family characteristics were taken into account.²³

CONCLUSIONS AND POLICY IMPLICATIONS

This review of preschool research leads to several conclusions. First, the historical proj-

ects like Abecedarian or Perry Preschool programs involve much more intensive interventions than contemporary preschool programs. Moreover, they each tested a very small sample drawn from a single community. As such, their results should not be generalized to current policies and should not be used to make inferences about the types of preschool programs being operated today.

Second, although preschool programs evaluated by the most rigorous research designs show modest but statistically significant improvements during the preschool years, these gains fade as children move into the kindergarten and first grades. The fadeout might be more accurately described as “catch up,” because the cognitive growth that occurs for all children in the early elementary grades is far greater than the gains during the preschool years, so it may be that children who did not have preschool simply caught up with those who did.

Third, all of the studies finding very large pre-K effects, particularly those in Tulsa, Boston, and Georgia, used regression discontinuity designs. These are not experimental designs, and their validity depends on demonstrating that the treatment and control groups—those having completed pre-K and those just starting pre-K, respectively—are close to identical but for the treatment. That evidence is missing in several respects, particularly attrition from the treatment group and time of testing. Most important, RDD studies cannot conduct longer-term follow ups because both the treatment and control groups have had preschool.

While the Perry Preschool results might not generalize to present-day pre-K programs, it is worth mentioning a viewpoint articulated by James Heckman and his colleagues about the potential role of preschool in changing future behavior. Their research leads to a sophisticated model that posits benefits of early intervention programs (and also adolescent intervention) leading to desirable adult behaviors, particularly high-school graduation, employment, and reduced criminality.²⁴ This suggests a need for more, longer-term studies

of children from preschool programs to determine adult behavioral changes.

Before considering expanding preschool offerings—especially making it universal—policymakers need to seek more randomized trials that track control and treatment groups over several years, and they should only attempt to replicate programs that have statistically significant, lasting effects, which can be achieved at scale and an affordable price. The original logic of helping disadvantaged children catch up may still be valid, because Head Start-type programs do show modest benefits during the preschool year. However, the proposal to expand preschool to everyone defeats the purpose of closing achievement gaps by giving disadvantaged children a “head start.” More importantly, the evidence as it currently exists demonstrates only short-term skill gains that fade after a few years, and there is insufficient evidence—two small, old, very intensive programs—for the type of long-term behavioral changes envisioned by the Heckman group.

Pre-K education may help, but the research to date does not support expanding existing government programs. New preschool programs should not be introduced unless they have statistically significant, non-negligible benefits.

NOTES

The author wishes to thank Grover “Russ” Whitehurst, Mike Puma, Nick Zill, and Chester E. Finn, Jr., for their helpful comments on earlier drafts of this paper.

1. A good example of this selective bias is found in the report by Hirokazu Yoshikawa et al., “Investing in Our Future: The Evidence Base on Preschool Education,” Society for Research in Child Development, October 2013.
2. For a review of the Abecedarian Project, see F. A. Campbell et al., “The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment,” *Developmental Psychology* 37 (2001): 231–42; for Perry Preschool, see Lawrence J. Schweinhart, Helen V. Barnes, and David P. Weikart, *Significant Benefits: The High/Scope Perry Preschool Study through Age 27* (Ypsilanti, MI: High Scope Press, 1993). The Chicago Parent-Child program is explained and evaluated in Arthur J. Reynolds, “Effects of a Preschool Plus Follow-On Intervention for Children at Risk,” *Developmental Psychology* 30 (1994): 787–804.
3. James J. Heckman, et al., “The rate of return to the High Scope Perry Preschool Program,” *Journal of Public Economics* 94 (2010): 114–28.
4. This study relies on two evaluations of Head Start published by the Department of Health and Human Services: Michael Puma, et al., “Head Start Impact Study, Final Report,” 2010, and Michael Puma et al., “Third Grade Follow-up to the Head Start Impact Study,” OPRE Report 2012-45, 2012.
5. A more detailed discussion of the HSIS findings can be found in David J. Armor and Sonia Sousa, “The Dubious Promise of Universal Preschool,” *National Affairs* no. 18 (Winter 2014).
6. See, for example, W. Steven Barnett et al., “Abbott Preschool Program Longitudinal Effects Study: Fifth Grade Follow-Up,” National Institute for Early Education Research, March 2013, nieer.org/publications/latest-research/abbott-preschool-program-longitudinal-effects-study-fifth-grade-follow.
7. Peter M. Bernardy, “Head Start: Assessing Common Explanations for the Apparent Disappearance of Initial Positive Effects” (PhD diss., George Mason University, 2012). This study used propensity analysis to produce equivalent groups of Head Start and control students.
8. See Mark W. Lipsey, et al., “Evaluation of the Tennessee Voluntary Prekindergarten Program: Kindergarten and First Grade Follow-Up Results from the Randomized Control Design,” Peabody Research Institute, Vanderbilt University, 2013.

9. Lipsey, et al., p. 12.
10. This review relies on Cheri A. Vogel, et al., “Early Head Start Children in Grade 5: Long-Term Follow-Up of the Early Head Start Research and Evaluation Study Sample.” OPRE Report # 2011-8, U.S. Department of Health and Human Services, 2010.
11. Nick Zill, “Ten Ideas for Improving Early Head Start—and Why the Program Needs Them,” in *Investing in Young Children: New Directions in Federal Preschool and Early Childhood Policy*, ed. Ron Haskins and W. Steven Barnett (Washington: Brookings Institution Center on Children and Families), pp. 39–48.
12. For the evaluations relied upon in this report, see the following: New Jersey: Ellen Frede et al., “The Abbott Preschool Program Longitudinal Effects Study (APPLES),” National Institute for Early Education Research, June 2007, <http://nieer.org/resources/research/APPLES.pdf>; Tulsa: William Gormley et al., “The Effects of Universal Pre-K on Cognitive Development,” *Developmental Psychology* 41 (2005): 872–84; Boston: Christina Weiland and Hirokazu Yoshikawa, “Impacts of a Prekindergarten Program on Children’s Mathematics, Language, Literacy, Executive Function, and Emotional Skills” *Child Development* 84 (2013): 2112–30.
13. See Peter Z. Schochet et al., “What Works Clearinghouse: Standards for Regression Discontinuity Designs,” June 2010, ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf. Also see Howard S. Bloom “Modern Regression Discontinuity Analysis,” MDRC Working Papers on Research Methodology, December 2009, www.mdrc.org/sites/default/files/full_446.pdf.
14. Some RDD evaluations follow a sample of children who did not take preschool, but this is not the control group in the RDD design. By being a self-selected group that chose not to enroll in preschool, it can differ from the treatment group in many ways, including motivation and other unmeasured characteristics (including initial skill levels).
15. These effects assume each school year consists of 10 months of learning.
16. There have been later evaluations, but this evaluation has more descriptive information available.
17. See Vivian C. Wong et al., “An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs,” *Journal of Policy Analysis and Management* 27, no. 1 (2008): 122–54.
18. Ellen S. Peisner-Feinberg et al., “Effects of Georgia’s Pre-K Program on Children’s School Readiness Skills: Findings from the 2012–2013 Evaluation Study,” Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill, March 2014.
19. Greg J. Duncan and Katherine Magnuson, “Investing in Preschool Programs,” *Journal of Economic Perspectives* 27, no. 2 (2013): 109–32.
20. Tennessee Department of Education, “2013–14 Tennessee Voluntary Pre-kindergarten Fact Sheet,” www.tn.gov/education/doc_earlylearning_prek/Pre-K_Fact_Sheet.pdf.
21. Chester E. Finn Jr., *Reroute the Preschool Juggernaut* (Stanford, CA: Hoover Institution Press, 2009).
22. Diane M. Early et al., “Teachers’ Education, Classroom Quality, and Young Children’s Academic Skills: Results from Seven Studies of Preschool Programs,” *Child Development* 78, no. 2 (2007): 558–80.
23. Terri J. Sabol and Robert C. Pianta, “Do Standard Measures of Preschool Quality Used in Statewide Policy Predict School Readiness?” *Education Finance and Policy* 9, no. 2 (2014): 116–64.
24. Flavio Cunha and James Heckman, “The Technology of Skill Formation,” *The American Economic Review* 97, no. 2 (2007): 31–47.

RELATED PUBLICATIONS FROM THE CATO INSTITUTE

The Poverty of Preschool Promises: Saving Children and Money with the Early Education Tax Credit by Adam Schaeffer, Cato Institute Policy Analysis no. 641 (August 3, 2009)

Universal Preschool Is No Golden Ticket: Why Government Should Not Enter the Preschool Business by Darcy Olson, Cato Institute Policy Analysis no. 333 (February 9, 1999)

The Advancing Nanny State: Why the Government Should Stay Out of Child Care by Darcy Olson, Cato Institute Policy Analysis no. 285 (February 9, 1997)

Caveat Emptor: The Head Start Scam by John Hood, Cato Institute Policy Analysis no. 187 (December 18, 1992)

RECENT STUDIES IN THE CATO INSTITUTE POLICY ANALYSIS SERIES

- 759. **Public Attitudes toward Federalism: The Public's Preference for Renewed Federalism** by John Samples and Emily Ekins (September 23, 2014)
- 758. **Policy Implications of Autonomous Vehicles** by Randal O'Toole (September 18, 2014)
- 757. **Opening the Skies: Put Free Trade in Airline Services on the Transatlantic Trade Agenda** by Kenneth Button (September 15, 2014)
- 756. **The Export-Import Bank and Its Victims: Which Industries and States Bear the Brunt?** by Daniel Ikenson (September 10, 2014)
- 755. **Responsible Counterterrorism Policy** by John Mueller and Mark G. Stewart (September 10, 2014)
- 754. **Math Gone Mad: Regulatory Risk Modeling by the Federal Reserve** by Kevin Dowd (September 3, 2014)
- 753. **The Dead Hand of Socialism: State Ownership in the Arab World** by Dalibor Rohac (August 25, 2014)
- 752. **Rapid Bus: A Low-Cost, High-Capacity Transit System for Major Urban Areas** by Randal O'Toole (July 30, 2014)
- 751. **Libertarianism and Federalism** by Ilya Somin (June 30, 2014)
- 750. **The Worst of Both: The Rise of High-Cost, Low-Capacity Rail Transit** by Randal O'Toole (June 3, 2014)

749. **REAL ID: State-by-State Update** by Jim Harper (March 12, 2014)
748. **State-Based Visas: A Federalist Approach to Reforming U.S. Immigration Policy** by Brandon Fuller and Sean Rust (April 23, 2014)
747. **Run, Run, Run: Was the Financial Crisis Panic over Institution Runs Justified?** by Vern McKinley (April 10, 2014)
746. **State Education Trends: Academic Performance and Spending over the Past 40 Years** by Andrew J. Coulson (March 18, 2014)
745. **Obamacare: What We Know Now** by Michael Tanner (January 27, 2014)
744. **How States Talk Back to Washington and Strengthen American Federalism** by John Dinan (December 3, 2013)
743. **The New Autarky? How U.S. and UK Domestic and Foreign Banking Proposals Threaten Global Growth** by Louise C. Bennetts and Arthur S. Long (November 21, 2013)
742. **Privatizing the Transportation Security Administration** by Chris Edwards (November 19, 2013)
741. **Solving Egypt's Subsidy Problem** by Dalibor Rohac (November 6, 2013)
740. **Reducing Livability: How Sustainability Planning Threatens the American Dream** by Randal O'Toole (October 28, 2013)
739. **Antitrust Enforcement in the Obama Administration's First Term: A Regulatory Approach** by William F. Shughart II and Diana W. Thomas (October 22, 2013)
738. **SNAP Failure: The Food Stamp Program Needs Reform** by Michael Tanner (October 16, 2013)
737. **Why Growth Is Getting Harder** by Brink Lindsey (October 8, 2013)
736. **The Terrorism Risk Insurance Act: Time to End the Corporate Welfare** by Robert J. Rhee (September 10, 2013)
735. **Reversing Worrisome Trends: How to Attract and Retain Investment in a Competitive Global Economy** by Daniel Ikenson (August 22, 2013)



Published by the Cato Institute, Policy Analysis is a regular series evaluating government policies and offering proposals for reform. Nothing in Policy Analysis should be construed as necessarily reflecting the views of the Cato Institute or as an attempt to aid or hinder the passage of any bill before Congress. Contact the Cato Institute for reprint permission. All policy studies can be viewed online at www.cato.org. Additional printed copies of Cato Institute Policy Analysis are \$6.00 each (\$3.00 each for five or more). To order, call toll free (800) 767-1241 or email catostore@cato.org.