From Preschool to Prosperity

Bartik, Timothy J.

Published by W.E. Upjohn Institute

Bartik, Timothy J.
From Preschool to Prosperity: The Economic Payoff to Early Childhood Education.
W.E. Upjohn Institute, 2014.
Project MUSE. muse.jhu.edu/book/82033.

For additional information about this book
https://muse.jhu.edu/book/82033

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2791977
Chapter 2

What Is the Evidence on the Earnings Effects of High-Quality Early Childhood Education, and Why Should We Believe It?

The empirical evidence is that high-quality early childhood education has large effects on a child’s future adult earnings. Early childhood education can increase a child’s future adult earnings by over 25 percent, and many programs have average future earnings effects of 3 percent or much greater. Over a career, even a 3 percent earnings boost amounts to many thousands of dollars.

But why should we believe this evidence? This chapter addresses this question.

WHY THE RESEARCH EVIDENCE FOR EARLY CHILDHOOD EDUCATION IS CREDIBLE

Why are the large benefit estimates for some early childhood programs believable? These estimates are believable because they come from rigorous research. Rigorous research means that the studies have good comparison groups. These comparison groups are made up of children similar to the children who participate in early childhood education, except for that participation. This comparability means that the estimated earnings increases are due to the program, not to preexisting differences between the program group and the comparison group.

Good comparison groups are needed because children are diverse. The child who participates in an early childhood education program may differ from the child who does not. Differences in observed child characteristics can be controlled for. We cannot
control for differences in child characteristics that are unobserved. If we lack good comparison groups, post-program differences may be due to unobserved characteristics, not the program. Perhaps program children have more ambitious parents, which biases the study toward finding positive program effects. Alternatively, perhaps program children have more preexisting problems, which lead the parents to enroll their children in the program. These preexisting problems will bias the research toward finding negative program effects.

The problem of bad comparison groups is called selection bias. Families self-select into enrolling their children in early childhood education. The families who self-select may have children with different unobserved characteristics from those who do not.

Bias may also occur because of the program’s selection procedures. The program may screen out children with too many problems, or it may enroll the neediest children. Outcome differences between the program and comparison groups may be due to this program selection.

RANDOM ASSIGNMENT STUDIES

How can research on early childhood education solve the selection bias problem? In some studies, the solution is that children are selected for the program using random assignment. Long-term effects of early childhood education are estimated in several well-done random assignment studies: Perry Preschool, the Abecedarian program, and the Nurse Family Partnership.

Perry was a half-day pre-K program at ages three and four in Ypsilanti, Michigan, from 1962 to 1967. The Abecedarian program was a full-time, full-year child care and pre-K program, from birth to age five, conducted from 1972 to 1977 in Chapel Hill, North Carolina. (The current Educare program is similar to Abecedarian.) The Nurse Family Partnership provides 2.5 years of nurse home visits to first-time mothers, from the prenatal period to age two. It has been subject to three experiments: 1) Elmira, New York (1977); 2) Mem-
What Is the Evidence and Why Should We Believe It? 11

...phasis (1987); and 3) Denver (1994). The NFP is an ongoing program with many U.S. sites. Perry, Abecedarian, and the NFP were all targeted at disadvantaged families.

These experiments found large long-term benefits. Perry Preschool on average increased its child participants’ future earnings by 19 percent. The Abecedarian program increased adult educational attainment and employment rates: The adult education effects predict a lifetime earnings increase of 15 percent. The higher adult employment rates, which exceed the boost expected from the educational attainment gains, bring the predicted lifetime earnings increase to 26 percent.5 The Nurse Family Partnership increased school test scores and reduced crime. These effects predict an average increase in lifetime earnings of 3 percent.6

Random assignment is the gold standard for ensuring that estimated program effects are credible. Random assignment means that we would expect the “treatment” and “control” groups to have the same average level of unobserved characteristics. (The “treatment” group are the children randomly assigned to participate in the program; the “control” group are the children randomly assigned to not participate.) Any sizable difference in outcomes between the treatment and control groups is likely due to the “treatment” (the program participation), not to unobserved characteristics.7

OTHER LONG-TERM STUDIES WITH GOOD COMPARISON GROUPS: HEAD START

In addition to these small random assignment experiments, rigorous evidence for long-run effects of early childhood education comes from large-scale programs that are not random assignment experiments but that still have good comparison groups.8 These studies have good comparison groups because they use data from natural experiments: natural accidents have resulted in similar persons having different access to early childhood education. Rigorous evidence for the long-term benefits of large-scale programs comes from studies of Head Start and of the Chicago Child-Parent Center Program.
Evidence on Head Start’s long-term effects comes from two types of studies with good comparison groups. First, a study by Ludwig and Miller (2007) compares long-run outcomes for children in two different types of counties: 1) counties that adopted Head Start in its early years and 2) otherwise similar counties that did not. Ludwig and Miller focus on counties whose early participation in Head Start occurred because Head Start provided extra technical assistance to high-poverty counties to help them write a good Head Start application. This technical assistance boosted Head Start program activity in these counties, compared to similar counties that just missed the poverty cutoff.

Ludwig and Miller compare long-term outcomes for children in counties that just made the cutoff for Head Start technical assistance versus children in similar counties that just missed the cutoff. This is a good comparison group, as it is hard to think of any reason, other than the Head Start participation, that would lead to large differences in child outcomes in counties on either side of the cutoff. Ludwig and Miller found that children in counties that made the cutoff had lower mortality rates after Head Start was begun, and higher educational attainment later on.

Second, other Head Start studies compare siblings, one of whom participated in Head Start and the other of whom did not. This is a good comparison group because the comparison holds constant unobserved family characteristics.

A sibling comparison study by Garces, Thomas, and Currie (2002) suggests that Head Start increases educational attainment and reduces crime for some groups. Another sibling comparison study by Deming (2009) finds effects of Head Start on adult outcomes that predict an average earnings increase of 11 percent.

THE CHICAGO CHILD-PARENT CENTER PROGRAM

Evidence on the long-term effects of large-scale pre-K programs comes from studies of the Chicago Child-Parent Center Program
(CPC). This program, started in 1967 and run by the Chicago Public Schools, provides participants with from one to two years of half-day pre-K at ages three and four. About half of the children in CPC participated only at age four, and the other half at both ages three and four.

Research on the CPC program uses as a comparison group children in similar neighborhoods in which the CPC program was not offered. This is a good comparison group because program participation depends on the accidents of geography, not on self-selection or program selection.9

CPC research has now followed former CPC participants, and the comparison group, through age 28. These results predict that CPC on average increases earnings by 8 percent.10

**PREDICTING LONG-TERM EARNINGS EFFECTS FROM SHORT-TERM TEST SCORE EFFECTS**

Direct evidence on the long-term earnings effects of early childhood education is better than indirect evidence. But long-term studies are scarce and expensive. Long-term studies also cannot produce results for more recent programs.

Long-term earnings effects can be predicted from a program’s effects on test scores in the short term. Research by Chetty et al. (2011), Currie and Thomas (2012), and others has produced good evidence on how early test scores affect adult earnings.11

How good are such predictions? For early childhood education, early test-score effects tend to underpredict long-run earnings effects. Table 2.1 shows test score predictions and adult earnings effects for four early childhood programs.12

The initial post-program test scores tend to underpredict eventual adult earnings effects.13 The third-grade test scores are even more of an underprediction of adult earnings effects. Possible reasons for this fading of results are discussed in Chapter 4. The bottom line is that test score predictions of adult earnings effects provide a conservative estimate of program benefits.
Table 2.1  Adult Earnings Effects, Compared with Predicted Earnings Effects Based on Test Scores

<table>
<thead>
<tr>
<th>Program</th>
<th>Long-run earnings effects based on adult outcomes (% of adult earnings)</th>
<th>Predicted earnings effects based on end-of-program test scores (% of adult earnings)</th>
<th>Predicted earnings effects based on third-grade test scores (% of adult earnings)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perry Preschool</td>
<td>19</td>
<td>12</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Abecedarian</td>
<td>26</td>
<td>13</td>
<td>7 to 11</td>
</tr>
<tr>
<td>Chicago Child-Parent Center Pre-K</td>
<td>8</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Head Start</td>
<td>11</td>
<td>3 to 6</td>
<td>1 to 3</td>
</tr>
</tbody>
</table>

NOTE: Adult earnings effects are shown as predicted average percentage increase in earnings due to the program, compared to expected earnings if the person had not participated in the program. End-of-program test scores are for the end of preschool or for kindergarten.

SOURCE: Author’s calculations, as explained in text and endnotes.
Hundreds of studies of early childhood education programs have estimated test score effects. Analyses of this research have been done by Camilli et al. (2010) and Leak et al. (2010). Leak et al. find average short-term test score effects that predict adult earnings effects of 5 percent.\textsuperscript{14} Camilli et al. find average test score effects that predict adult earnings effects of 9 percent.\textsuperscript{15}

But the Camilli et al. and Leak et al. meta-analyses are not precise enough to guide policy. The average effects come from studies with diverse methodologies of variable quality, and the programs studied have different designs, costs, and quality.

This book will focus on a few of these studies that have particularly good comparison groups. These studies include the following three categories: 1) the Infant Health and Development Program; 2) North Carolina’s Smart Start and More at Four programs; and 3) studies of state and local pre-K programs that have been done using regression discontinuity methods, which will be explained below.

**INFANT HEALTH AND DEVELOPMENT PROGRAM**

The Infant Health and Development Program (IHDP) was an experiment conducted from 1985 to 1988 at eight U.S. sites. The program provided high-quality full-time child care at ages one and two.\textsuperscript{16} The program replicated features of the Abecedarian program. However, whereas the Abecedarian program provided full-time child care and pre-K from birth to age five, the IHDP only provided such services for two years.

The IHDP was a random assignment experiment. Duncan and Sojourner’s (2013) analysis of these data suggest that for low-income children, the IHDP had many statistically significant and substantively large effects on test scores.\textsuperscript{17} The third-grade test effects predict that the IHDP would on average boost adult earnings by 12 percent.\textsuperscript{18}
NORTH CAROLINA’S SMART START PROGRAM, 
AND MORE AT FOUR

Smart Start is a state of North Carolina program, begun in 1993, that provides state aid to county partnerships that try to improve the quality of local early-childhood services, from birth to age four. Most Smart Start funds have been devoted to child care, including child care subsidies to low-income families and training for child care centers.

More at Four was a state of North Carolina program from 2001 to 2011 that supported full-day pre-K at age four. It has since been renamed NC Pre-K.

A recent study (Ladd, Muschkin, and Dodge 2014) relies on a natural experiment: Both Smart Start and More at Four were gradually rolled out to different counties. The researchers use this natural variation in funding by county and time period to estimate these programs’ effects on third-grade test scores. This study has a good comparison group because it is unclear why third-grade test scores in different counties would change in response to lagged program funding except for a true program effect.

Based on Ladd et al.’s estimated test score effects, typical levels of Smart Start or More at Four funding are predicted to increase average future earnings in a county by 1.6 percent for Smart Start and by 2.9 percent for More at Four. Over an entire career, a 1.6 percent earnings boost is a lot of money.

These average effects include many children who receive few or no services from these two programs. In the typical county, only 25 percent of four-year-olds participated in pre-K funded by More at Four, so it is remarkable that the program increases average earnings by 2.9 percent. Extrapolation implies that increasing a county’s enrollment in More at Four from zero to 100 percent would boost earnings by over 11 percent. For Smart Start, average spending per child under the age of five is only $220 annually, so the 1.6 percent earnings effect is large compared to this cost.
STATE AND LOCAL PRE-K PROGRAMS:
THE REGRESSION DISCONTINUITY EVIDENCE

Rigorous evidence for large-scale state or local pre-K programs comes from studies that look at student performance on the same test at entrance to the pre-K program for four-year-olds and at entrance to kindergarten. Such studies have been done in many states, including New Jersey, Michigan, West Virginia, South Carolina, Oklahoma, New Mexico, and Tennessee, and in cities such as Tulsa, Boston, and Kalamazoo (Bartik 2013; Bartik, Gormley, and Adelstein 2012; Hustedt, Barnett, and Jung 2008; Hustedt et al. 2010; Weiland and Yoshikawa 2013; Wong et al. 2008).

These studies use a methodology called “regression discontinuity.” This methodology relies on pre-K and kindergarten entrance being based on an age cutoff. Students entering pre-K, and students who graduated from pre-K and are entering kindergarten, are similar in family or program factors that led to pre-K participation. Selection bias should not be a big problem.

What does explain why one child is a pre-K entrant, and the other child is a pre-K graduate, is the child’s age. The child entering kindergarten will on average be one year older than the child entering pre-K. But ages vary over a one-year period for each group. There are children entering pre-K who are just a few days younger than children entering kindergarten. These children should be similar in both observed and unobserved characteristics, except that the slightly older child has benefited from a year in a pre-K program.

We can estimate how much higher student test scores are at kindergarten entrance for those children who went to preschool, compared to children just starting in those same preschool programs. Using the variation in test scores with age, we can estimate how test scores vary with age. We can then see whether there is a “jump”—a “discontinuity”—in scores for the child who is just old enough for kindergarten entrance, and who has experienced a year of pre-K, versus the child who is a little too young for kindergarten entrance, and therefore is just entering pre-K.
Figure 2.1 shows an example of this jump in student test scores associated with Tulsa’s pre-K program (Bartik, Gormley, and Adelstein 2012). The figure shows fall test scores along the vertical axis for groups of Tulsa students, sorted along the horizontal axis by age.20 The students to the left of the vertical line are too young to enter kindergarten and are just entering Tulsa’s pre-K program. The students to the right of the vertical line are former Tulsa pre-K participants who are just entering kindergarten. All these students were given the same tests at the same time in the fall.

Test scores in most of the figure go up smoothly with age. But there is an abrupt jump at the cutoff between pre-K entrants and kindergarten entrants. The most plausible explanation of this jump is that Tulsa’s pre-K program increases test scores. The students just to the left and right of the cutoff are almost the same age, are shown in the full study to be similar in all observed characteristics, and differ mainly in that the group just to the right has had one year of Tulsa pre-K. Therefore, it seems likely that the jump is due to pre-K.21

Many studies have found such test score jumps in comparing entrants in state pre-K programs with kindergarten entrants who are pre-K graduates. These test score jumps tend to be in the range of increasing student learning during the pre-K year by perhaps 40 to 80 percent, compared to what children would learn on their own, without the state pre-K program.22

We can combine the estimated impact of state pre-K on kindergarten test scores, and the estimated impact of kindergarten scores on adult earnings, to predict how much these state pre-K programs increase adult earnings. These calculations suggest that state and local pre-K programs increase future adult earnings of children from low-income families by 6 to 15 percent.23

The study of Tulsa’s full-day pre-K program by Bartik, Gormley, and Adelstein (2012) estimates test score effects that imply average adult earnings effects, for children from low-income families, of 10 percent. A study of Boston’s full-day pre-K program by Weiland and Yoshikawa (2013) estimates test score effects that imply
average adult earnings effects, for children from low-income families, of 15 percent. 

SUMMARY OF THE EVIDENCE

We have better evidence for the effectiveness of early childhood education than for almost any social or educational intervention. We have better evidence because we have good comparison groups. Ironically, these good comparison groups arise because early childhood education is not universal, so many children are excluded from services.

Early childhood education has more evidence for effectiveness than most educational programs and practices. Consider the evidence for whether children benefit from going to third grade. No evidence from random assignment or natural experiments shows that children
do any better from attending third grade rather than staying home. For third grade, we are unable to get good comparison groups, because everyone goes to third grade.

But how do these adult earnings benefits for high-quality early childhood education compare with program costs? The next chapter considers this important issue.
This page left blank.
Timmy's been selected for the control group. He can't attend 3rd grade.