



SCHOOL of  
PUBLIC POLICY

## 6

# Currie/Thomas Econometric Studies

**Douglas J. Besharov**  
**Peter Germanis**  
**Caeli A. Higney**  
**and**  
**Douglas M. Call**

September 2011



Maryland School of Public Policy  
Welfare Reform Academy  
[www.welfareacademy.org](http://www.welfareacademy.org)

Part of a forthcoming volume  
*Assessments of Twenty-Six Early Childhood Evaluations*  
by Douglas J. Besharov, Peter Germans, Caeli A. Higney, and Douglas M. Call

*Note:* This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to [welfareacademy@umd.edu](mailto:welfareacademy@umd.edu) or fill out the comment form at [http://www.welfareacademy.org/pubs/early\\_education/chapter6.html](http://www.welfareacademy.org/pubs/early_education/chapter6.html).

## 6

# Currie/Thomas Econometric Studies

The federal Head Start program, started in 1965, seeks to enhance “the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families.”<sup>1</sup>

In a series of studies, University of California, Los Angeles economists Janet Currie, Duncan Thomas, and Elana Garces (the “UCLA team”) used econometric statistical models to analyze data from three large, nationally representative surveys to estimate the impact of Head Start. The surveys were the National Longitudinal Survey of Youth (NLSY), the Supplemental National Longitudinal Survey’s Child-Mother file (NLSCM), and the Panel Study of Income Dynamics (PSID). The UCLA team estimated the impact of Head Start on children who participated in the program by examining a range of short-term and long-term outcomes. The first study used data from 1986 and 1990, the second used data from 1979–1990, and the third used data from 1968–1977. The authors compared children who participated in Head Start to those who did not, statistically controlling for socioeconomic differences and other factors. They conclude that Head Start participation was associated with a number of positive cognitive effects for both white and black children, but that these gains diminished over time for black children. Long-term findings include improved school performance for white children and a reduction in criminal activity for black children.

The studies used large, nationally representative data sets; however, the methodology used (sibling comparison) resulted in a sample that was not nationally representative. In addition, there are a number of methodological problems, including (1) possible selection bias if the parents enrolled either the less able or the more able sibling in Head Start, (2) the possibility of parents making “compensating investments in the non-Head Start child;”<sup>2</sup> (3) spillover effects from the siblings in Head Start sharing what they learn with the sibling not in Head Start; (4) data quality questions in the NLCSM because, in different waves of the study, mothers gave different responses to the same question concerning their child’s Head Start participation; and (5) serious

---

<sup>1</sup>U.S. Department of Health and Human Services, Office of Head Start, “About the Office of Head Start,” <http://www.acf.hhs.gov/programs/ohs/about/index.html> (accessed May 13, 2010).

<sup>2</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 344.

non-random attrition in both the PSID and NLSY. Thus, these findings are subject to considerable uncertainty, especially because they find long-term impacts for white children and not for black children, even though most experimental studies find positive impacts for black children. Hence, these studies should be viewed as suggestive only.

## Program Design

**Program group.** From its inception, Head Start has been targeted to children with family incomes below the poverty line at the time of enrollment.<sup>3</sup> The UCLA team excluded children who participated in the earliest years of Head Start, when it was predominantly a summer program.

Two of the UCLA team's studies reported the findings separately for whites and blacks and a third study reported findings for Hispanics. On average, mothers whose children participated in Head Start were more disadvantaged than mothers whose children did not. For example, in one study examining the early cognitive effects of Head Start, both white and black Head Start children came from families with lower incomes than children who did not attend preschool.<sup>4</sup> In addition, both white and black Head Start children had mothers and grandmothers with less education and who scored lower on the Armed Forces Qualifying Test (AFQT), considered a reasonable test of reading comprehension, vocabulary, and mathematics skills, although the gaps were substantially larger among whites.

**Services.** The services provided to children by Head Start vary from site to site, as the program is implemented on the local level. The program as a whole provides a wide range of services geared at improving school readiness, including, but not limited to, education, health, and nutrition components.

**The Evaluation.** In their first study, the UCLA team examined a sample of children born to women participating in the National Longitudinal Survey of Youth (NLSY).<sup>5</sup> The NLSY is an annual survey that began in 1979 when the participants were between the ages of fourteen and twenty-one. As of the 1990 wave, these women were ages twenty-five to thirty-two and had given birth to 8,500 children. In 1986, the survey was supplemented with the biannual National Longitudinal Survey Child-Mother (NLSCM), a separate survey of the children of female NLSY

---

<sup>3</sup>Ten percent of families may have initial incomes above the poverty line

<sup>4</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 349.

<sup>5</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 341–364.

participants. If old enough, the children received a number of standardized tests, including the Peabody Picture Vocabulary Test (PPVT) and the Peabody Individual Achievement Tests in Reading Recognition and Mathematics (PIAT-Reading and PIAT-Math). The mothers were also asked if their children ever attended Head Start or some other preschool (in the 1988 and 1990 waves).

Using data from the 1990 NLSY and the 1986 NLSCM, the UCLA team used regression analysis to estimate the effects of Head Start participation on PPVT scores, grade retention, and immunization status. They controlled for a range of observable mother and child characteristics, including child's age, sex, and whether the child was the first born; a measure of household permanent income; the mother's education, AFQT score, height, and number of siblings when she was fourteen years old; and the maternal grandmother's educational attainment. The UCLA team recognized, however, that because Head Start was voluntary, mothers could choose whether to enroll their children in the program. To control for unobserved characteristics associated with selection into the program, they adopted a fixed-effects model.<sup>6</sup> As the UCLA team explains:

These models control for constant characteristics of households including permanent income, maternal education, and other measures of (unobserved) family background and tastes. If it is primarily these constant factors that determine participation in Head Start, then fixed-effects models will provide unbiased estimates of the true program effects.<sup>7</sup>

Their fixed-effects estimates of Head Start were based on within-family comparisons of siblings who attended Head Start to those who did not.<sup>8</sup>

In the second study, the UCLA team examined the effects of Head Start on Hispanic children (where at least one parent was of Latino descent).<sup>9</sup> They used data from the NLSY (emphasizing PPVT and PIAT scores) through 1992, when the women were between the ages of twenty-seven and thirty-four. In this study, they restricted the analysis to children age five and older at the time of the survey, because they had completed Head Start.

In the third study, published in 2002, the UCLA team used the PSID to examine the long-

---

<sup>6</sup>The fixed-effects model included fewer control variables: child age, sex, and whether the child was the first born, as well as household income when the child was three. Most of the other maternal variables included in the regression were dropped, because they were captured by using the fixed-effects model.

<sup>7</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 343.

<sup>8</sup>The analysis was restricted to Head Start children and siblings age three and older.

<sup>9</sup>Janet Currie and Duncan Thomas, "Does Head Start Help Hispanic Children?" *Journal of Public Economics* 74, no. 2 (November 1999): 235–262.

term effects of Head Start participation for a sample of persons born between 1964 and 1977—examining outcomes that included high school completion, college attendance, earnings, and criminal behavior.<sup>10</sup> The PSID began in 1968 with a survey of 4,800 households. In 1995, a one-time special supplement to the survey provided additional information about early childhood experiences.

## Major Findings

The studies conducted by the UCLA team suggest that Head Start participation was associated with a number of positive cognitive effects for white, black, and Hispanic children, but that these gains “faded out” for black children.<sup>11</sup> They also reported that the program improved school performance for white children and reduced criminal activity for black children when they were young adults.

**Cognitive.** In their 1995 paper, the UCLA team estimated the effects of Head Start separately for whites and blacks. Scores from the PPVT were used to measure academic performance for children age four and older. Outcomes were expressed as percentile scores based on nationally accepted norms for the age and sex of the child. Although white children who attended Head Start had lower PPVT scores than those who did not, there was no statistically significant difference among white children once the UCLA team controlled for observable characteristics. But, when they controlled for *unobserved* differences between families using the fixed-effects model, Head Start participation was estimated to increase PPVT scores by 6 percentile points closer to the national norm. They found no statistically significant effects for black children of all ages regardless of the methodology used.

The UCLA team also estimated the persistence of these effects by examining how the effects varied by the age of the child. They found that, when measured immediately at the end of the program, both white and black children who participated in Head Start had statistically significant gains of nearly 7 percentile points on the PPVT. By age ten, however, while the gains for whites persisted (about 5 percentile points), the gains for blacks were no longer evident.<sup>12</sup> (The UCLA team noted that in an earlier paper they had reported results using the PIAT-Math

---

<sup>10</sup>Eliana Garces, Duncan Thomas, and Janet Currie, “Longer-Term Effects of Head Start,” *American Economic Review* 92, no. 4 (September 2002): 999–1012.

<sup>11</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 341–364; Janet Currie and Duncan Thomas, “Does Head Start Help Hispanic Children?” *Journal of Public Economics* 74, no. 2 (November 1999): 235–262; and Eliana Garces, Duncan Thomas, and Janet Currie, “Longer-Term Effects of Head Start,” *American Economic Review* 92, no. 4 (September 2002): 999–1012.

<sup>12</sup>The number of children age ten and older was smaller than (and a subset of) the number of children five and older. Although the apparent difference in impacts by age could be due to “fade out” for black children, it could also be due to differences in the characteristics of children in the two groups.

and PIAT-Reading test scores. The results for reading recognition were similar, but weaker, and the math impacts were not statistically significant.<sup>13</sup>)

For Hispanic children, in their 1999 paper, the UCLA team measured academic performance using the PPVT, and PIAT-Math and PIAT-Reading scores. Outcomes were measured for children age five and older and were expressed as percentile scores based on nationally accepted norms for the age and sex of the child. Although Hispanic children who attended Head Start had lower scores than those who attended another preschool, there were no statistically significant differences among the children once the UCLA team controlled for observable characteristics. When they controlled for *unobserved* differences using the fixed-effects approach, Head Start participation was found to increase PPVT scores by 10 percentile points. Similarly there was a 5 percentile point increase in the PIAT-Math scores. There were no statistically significant findings with respect to PIAT-Reading scores. (There are no data for older Hispanic children.)

**School readiness/performance.** The UCLA team also assessed school performance, using grade retention, high school completion, and college attendance data.

*Grade retention.* The PSID asked mothers of children age ten and older if their children had repeated any grades. Analyzing these data, the UCLA team found that Head Start had no statistically significant effects on the probability of grade retention when observable characteristics were controlled for. The fixed-effects estimates, however, suggested that white children who participated in Head Start were 47 percent less likely to repeat a grade than their non-Head Start siblings. There was no statistically significant effect for blacks. For Hispanic children, after controlling for observable characteristics, the UCLA team (1999) found that Head Start participation decreased the likelihood of repeating a grade by 10 percent. Controlling for unobserved fixed-effects increased the estimated effect to 22 percent.

*High school completion.* The UCLA team (2002) found that after controlling for observable characteristics,<sup>14</sup> Head Start attendance did not have a statistically significant impact on high school completion. Nor did they find a statistically significant difference when they added controls for maternal fixed-effects. Stratifying by race, however, they found that whites who went to Head Start were 20 percentage points more likely to finish high school than their siblings who did not attend Head Start. The impact for those whose mothers had no more than a high school education was 28 percentage points. There were no statistically significant findings

---

<sup>13</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 345.

<sup>14</sup>Characteristics included: year of birth, gender, race, maternal and paternal education, whether the mother was the head of the household, family incomes at age four to six, family size at age four, birth order, whether the respondent was the oldest child, and whether the respondent had been of low birthweight.

for blacks.

*College attendance.* The UCLA team (2002) found that after controlling for observable differences, those who had attended Head Start were 7.5 percentage points more likely to attend college. When they added controls for maternal fixed-effects, the impact became slightly larger, but they described the statistical significance as “marginal.” Stratifying by race once again resulted in large effects for whites. Whites who attended Head Start as children were 28 percentage points more likely to attend college. Again, there were no statistically significant effects for blacks.

**Socioemotional development.** Relevant tests apparently not administered or results not reported.

**Health.** The UCLA team reports that, after controlling for observable and unobservable differences, Head Start participation increased the probability of measles immunization by 8 to 9 percent for both white and black children. They found no impact on nutritional and health status as measured by height for age.

**Behavior.** Data apparently either not collected or not reported.

**Crime/delinquency.** The UCLA team (2002) examined whether Head Start participation affected self-reported criminal behavior, using a broad definition of criminal behavior that included major felonies as well as less serious offenses, such as drunk driving. They also included responses to a question about having been “booked or charged” with a crime, including, in some cases, those later found innocent. The researchers found no statistically significant effects of Head Start attendance for whites. For blacks, they found a statistically significant reduction in criminal behavior of nearly 12 percentage points.

**Early/nonmarital births.** Data apparently either not collected or not reported.

**Economic outcomes.** The UCLA team (2002) found no statistically significant effects on labor force participation or earnings among young adults overall for either race. They did report, however, a statistically significant increase in annual earnings for former Head Start participants who were white, in their early twenties, and whose mothers had not completed high school.

**Effects on parents.** Data apparently either not collected or not reported.

**Benefit-cost findings.** The UCLA team (1995) reports that, at the time of their study, Head Start cost approximately \$4,900 per child (in 2005 dollars). In their first study, they conclude that Head Start, “when viewed strictly in terms of lasting benefits,” is not cost effective for blacks, because the program’s cost greatly exceeds the estimated value of health care services

provided.<sup>15</sup> Elsewhere, they point out that there are many short-term benefits that should not be ignored in determining the program's benefit-cost ratio.<sup>16</sup> For whites, however, they argue that the gains are larger, "since even a small decline in the high school dropout rate has the potential to pay for itself in terms of future wage gains."<sup>17</sup> They do not provide any specific data to support this position. In any event, they base their conclusion on benefits to participants, rather than on net savings to taxpayers.

### Overall Assessment

The studies used large, nationally representative data sets, but because the children had not been randomly assigned to program or control groups, the findings may be subject to selection bias. In addition, various shortcomings in the data from the surveys (for example, inaccuracies about participation in Head Start) further undermine the confidence that can be placed in their findings. Although the UCLA team attempted to control for selection bias using sophisticated statistical models, such modeling remains controversial. Thus, these findings are subject to considerable uncertainty, especially since they find long-term impacts for white children and not for black children, even though most experimental studies find positive impacts for black children. Hence, these studies should be viewed as suggestive only.

**Program theory.** Apparently, there is no specific theory detailed beside the general expectation that early intervention programs promote school readiness and improve developmental outcomes for children.

**Program implementation.** The UCLA team noted that program implementation was uneven throughout the country and that this could be a potential explanation for differential effects for blacks and whites.

**Assessing the randomization.** The groups were not randomly assigned.

**Assessing statistical controls in experimental and nonexperimental evaluations.** The UCLA team used a fixed-effects model in an attempt to deal with potential selection bias,

---

<sup>15</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 361.

<sup>16</sup>Janet Currie, "Early Childhood Education Programs," *Journal of Economic Perspectives* 15, no. 2 (Spring 2001): 213–238.

<sup>17</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 361.

comparing siblings who participated in Head Start to those who did not.<sup>18</sup> The use of sibling models was intended to control for a range of constant household characteristics, including unmeasured characteristics. One problem with a fixed-effects model, however, is that the influence of family background may not remain constant over time. Moreover, the model, which excluded single children, relied on the assumption that children with siblings are not systematically different from children with no siblings.

The UCLA team acknowledges that a fixed-effects model could produce biased estimates if there were child-specific factors that affected participation. For example, if parents enrolled the least able sibling (perhaps because they felt the child would gain the most), Head Start's effects may be understated. On the other hand, if parents enrolled the more able sibling, the program's effects may be overstated. The UCLA team examined some observable characteristics surrounding Head Start entry and conclude that, "if anything, Head Start children are worse off than non-Head Start siblings."<sup>19</sup> Thus, they argued that their estimates were conservative. Of course, unmeasured differences could work in the opposite direction, so this issue cannot be resolved. Greg Duncan and Christina Gibson-Davis write, "If Head Start use differs between siblings because of family events that may themselves have detrimental effects on the children (e.g., an eviction forced the family to move away from a service area), then it is more likely that sibling models estimates are biased."<sup>20</sup> And Steven Barnett, director of the National Institute for Early Education Research, and Gregory Camilli, senior research fellow at the National Institute for Early Education Research, point out that providing Head Start to one child could lead to "families attempting to compensate children who did not attend Head Start for this lost opportunity."<sup>21</sup>

In addition, effects may be underestimated if there were "spillover" effects from one sibling to another. As the UCLA team explains: "Spillover effects can be important because a child teaches his or her sibling something learned in Head Start, because the parent gains access to a service that is of benefit to both children, or because the parent makes compensating

---

<sup>18</sup>There are other approaches for dealing with selection bias. For example, researchers have used instrumental variables estimators to estimate the impact of other social programs. The UCLA team indicates that they experimented with this approach but could not find "convincing" instruments.

<sup>19</sup>Janet Currie and Duncan Thomas, "Do the Benefits of Early Childhood Education Last?" *Policy Options* 18, no. 6 (July/August 1997): 48.

<sup>20</sup>Greg J. Duncan and Christina M. Gibson-Davis, "Connecting Child Care Quality to Child Outcomes: Drawing Policy Lessons from Nonexperimental Data" *Evaluation Review* 30:622, <http://erx.sagepub.com/cgi/reprint/30/5/611.pdf> (accessed May 23, 2008).

<sup>21</sup>W. Steven Barnett and Gregory Camilli, "Compensatory Preschool Education, Cognitive Development, and 'Race'" in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 395.

investments in the non-Head Start child.”<sup>22</sup> Barnett and Camilli echo this concern: “It is difficult to draw conclusions about effects of parents on children. However, theory and research in child development predicts that Head Start’s parent activities should have impacts on development and school success for all children in a family.”<sup>23</sup>

Finally, the UCLA team notes that participant selection is often made by program administrators and that estimates that do not take this selection process into account may be biased.<sup>24</sup>

Thus, despite the careful analysis conducted by the UCLA team, there is always some uncertainty regarding the findings of “quasi-experiments.” Mark Lipsey and David Cordray of Vanderbilt University argue that the existing techniques to control for selection bias do not resolve the problem completely: “The major problem with the available statistical approaches to selection bias is the sensitivity of the results to the violation of the model assumptions, especially the requirements that all relevant variables be specified. Much work remains to be done on the question of which models are best for which circumstances.”<sup>25</sup>

**Sample size.** Although the sample sizes in each of the data sets used by the UCLA team were quite large, the number of children in sibling pairs where one participated in Head Start and one did not was relatively small. The UCLA team explained that the estimates of Head Start “effects are identified by within-family variation in enrollments, so we need a large enough sample of ‘changers’ to identify the effects.”<sup>26</sup> The number of “changers” turned out to be very small in each of the studies.

For example, the NLSY sample used for the UCLA team (1995) began with about 8,500 children, because the fixed-effects estimates of Head Start were based on “within family comparisons of children in Head Start with siblings who did not attend any preschool,” this further reduced the comparison to only 134 sibling pairs for white children and 94 sibling pairs

---

<sup>22</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 344.

<sup>23</sup>Steve Barnett and Gregory Camilli, “Definite Results from Loose Data: A Critique of ‘Does Head Start Make a Difference?’” (unpublished paper, October 16, 1997), 14.

<sup>24</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 343.

<sup>25</sup>Mark W. Lipsey and David S. Cordray, “Evaluation Methods for Social Intervention,” *Annual Review of Psychology* 51 (February 2000): 349.

<sup>26</sup>Janet Currie and Duncan Thomas, “Does Head Start Help Hispanic Children?” *Journal of Public Economics* 74, no. 2 (November 1999): 241.

for black children.<sup>27</sup>

For their sample of Hispanics, the UCLA team (1999) used the NLSY with data on 750 Latino children, of which 182 participated in Head Start. There were only sixty-five sibling pairs, however, in which one child participated in Head Start and the other did not.<sup>28</sup>

In their analysis of long-term outcomes, the UCLA team (2002) began with the PSID sample of 3,255, with 489 Head Start participants. The fixed-effects model reduced the sample size to 1,742, further subdivided by race, with 1,036 whites and 706 blacks. Of these, 255 respondents (from 100 families) were in families where at least one child participated in Head Start and one did not. The numbers for whites and blacks were smaller, and the number for subgroups within each of these, smaller yet.

Although the UCLA team acknowledges that the sample was small, they point out that many experimental evaluations have used even smaller samples. While this may be true, experimental projects can get away with smaller samples because concerns about the comparability of program and control groups are less serious. Moreover, numbers as small as these would be a concern for experimental evaluations as well. Nonetheless, Thomas argues that statistical significance when there is a small sample size is evidence of strong results.<sup>29</sup>

**Attrition.** Both the PSID and NLSY have had serious attrition over time. Over 30 percent of the NLSY sample was lost because the subject could not be found for the interview, failed to provide a valid answer to the Head Start question, or lacked a test score.<sup>30</sup> Steve Barnett and Gregory Camilli note, “This attrition was hardly random.”<sup>31</sup> As noted above, most attrition

---

<sup>27</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 349. For grade retention, the sample size is considerably smaller, because only families with siblings ten years old or older are included.

<sup>28</sup>The UCLA team did not provide the exact number. Instead, they simply indicated that there were 130 children in the category identified as “some children in Head Start and some with no preschool.” Janet Currie and Duncan Thomas, “Does Head Start Help Hispanic Children?” *Journal of Public Economics* 74, no. 2 (November 1999): 242.

<sup>29</sup>Duncan Thomas, University of California, Los Angeles, e-mail message to Peter Germanis, August 13, 2003.

<sup>30</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 392.

<sup>31</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 392.

occurred because the fixed-effects model employed by the UCLA team used only children in families where one child attended Head Start and the other did not. Barnett and Camilli found that the cognitive test scores of this group significantly differed from those of Head Start children who were the only child in their family and Head Start children whose siblings also attended the program.<sup>32</sup> Thus, this attrition resulted in a sample that is unrepresentative of the entire Head Start population.

**Data collection.** The data were from three large, nationally representative surveys: the NLSY, NLSCM, and PSID.

**Measurement issues.** Many outcomes and background characteristics were obtained from survey data, with some involving questions about Head Start or preschool participation many years earlier. It is unclear how well these surveys captured the actual participation and characteristics of Head Start families. Aware of this possibility, the UCLA team examined the PSID and found that the Head Start participation rates were similar to those derived from administrative data on Head Start participation. For the NLSY, they note, “The original NLSY oversampled the poor and so a relatively large proportion of the sample children—about one-fifth—participated in Head Start. In addition, due to oversampling there are large enough numbers of African-Americans to allow separate examination of this group.”<sup>33</sup>

The UCLA team does not, however, address the fact that Head Start participation may have been substantially overreported in the NLSCM.<sup>34</sup> Furthermore, rates of reported Head Start attendance were especially high among higher income African-American families and, according to Barnett and Camilli, “This suggests that the NLSCM sample is either highly unrepresentative or that erroneous reporting occurs frequently.”<sup>35</sup> For instance, within a group of children for

---

<sup>32</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 393.

<sup>33</sup>Janet Currie and Duncan Thomas, “Does Head Start Make a Difference?” *American Economic Review* 85, no. 3 (June 1995): 345.

<sup>34</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 393. They assert, “A second data limitation of the NLSCM sample is that the treatment indicator (whether a child attended Head Start or not) contains serious errors because it is based on retrospective parent report. Data from the National Household Education Survey suggest that parents may have difficulties distinguishing Head Start from other types of programs so that Head Start participation is substantially overreported.”

<sup>35</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 393.

whom parental reports of Head Start participation were collected in both 1988 and 1990, a number of children were reported in 1988 as having been in Head Start, but in 1990 were reported as never having been in Head Start.<sup>36</sup> Barnett and Camilli find, “As a percentage of children in 1990, the latter error rates for African-Americans, Latinos, and non-Latino whites were 16%, 21%, and 4%, respectively.”<sup>37</sup>

The UCLA team based the cognitive outcomes on standardized test scores, such as the PPVT or PIAT. In some cases, the different test scores seemed to suggest different findings. For example, in the 1995 paper, the PIAT findings were much weaker than the PPVT findings, but the UCLA team did not elaborate on the implications of these differences. Barnett and Camilli find these differences concerning:

The mean percentile scores on the PPVT are lower than those on the PIAT tests for all groups, but the gap in scores is much larger for minority children, and is most severe for African-American children. Inspection of the PPVT-R percentile data reveals that minimum (floor) scores were obtained much more frequently for African-American and Latino children than for white children. This problem was much less common with the PIAT scores. The comparisons . . . raise questions about the validity of the PPVT-R for minority children and about the use of percentile scores in statistical analyses.<sup>38</sup>

Barnett and Camilli conducted a reanalysis of the effects of Head Start on raw scores for the PIAT and PPVT-R. Using the UCLA team’s data, they averaged each child’s raw scores to produce a single score for each test and used a simple fixed-effect model to reanalyze the effects of Head Start. To test for “fade out,” they added the average age tested as a regressor in the fixed-effects model. They found that Head Start only had a positive, statistically significant impact on African-American children, with an effect that is roughly equivalent to four items on the PIAT reading recognition test. They also found evidence of “fade-out” for African American children. For white children, estimated effects on the PIAT-reading test were smaller than for African American and were not statistically significant. For Hispanic children, Barnett and Camilli found the “surprising result” that Head Start had no immediate effect on the PPVT-R, but that children showed increasingly large gains on the test as they became older. Assessing this result, Barnett

---

<sup>36</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 394.

<sup>37</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 394.

<sup>38</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 394.

and Camilli note, “No theory or previous study has suggested this pattern of effects for Latinos or any other children. In our view, the results are exactly the sort of nonsense one can expect when flaws plague both the data and the analytical model, and no conclusion should be drawn from any fixed-effect estimates based on the NLSCM data.”<sup>39</sup>

Turning to other variables in their analysis, Barnett and Camilli find that firstborn status is sometimes associated with more positive outcomes. They also find that measured differences in the home environment (as it relates to intellectual stimulation) strongly affect children’s test scores, which “suggests that unmeasured intrafamily differences among siblings may well bias the fixed-effects model’s Head Start estimates toward zero.”<sup>40</sup> Thus, they believe that the UCLA team’s analysis tends to understate Head Start’s effects.<sup>41</sup> Barnett and Camilli conclude, “The results of our reanalysis clearly differ from the [Currie and Thomas, 1995] results and are at least as plausible. One could argue the merits of each analysis, but it seems wiser to conclude that the NLSCM data are highly problematic for the purpose of evaluating the efficacy of Head Start.”<sup>42</sup>

**Generalizability.** Many of the outcomes are based on Head Start participation twenty to thirty-five years ago. The program and population have changed considerably since that time. Additionally, the studies’ samples are not necessarily representative of the entire Head Start population. According to Barnett and Camilli, for example, the NLSCM Head Start sample was less economically disadvantaged than the true Head Start population. Thus, the UCLA team’s findings are not necessarily generalizable to the entire Head Start population.

In addition, the studies only looked at Head Start children with siblings. The results, therefore, are not generalizable to Head Start children who have no siblings.

**Replication.** Although the UCLA team used three different data sets, the two studies could be considered replications; however, they examined different outcomes. Neither study has been replicated by other researchers.

**Evaluator’s description of findings.** In their published papers, the UCLA team did not provide estimates of the impact of Head Start overall, but examined outcomes for various racial/ethnic subgroups, as well as subgroup breakdowns by maternal education/AFQT and, for

---

<sup>39</sup>W. Steven Barnett and Gregory Camilli, “Compensatory Preschool Education, Cognitive Development, and ‘Race’” in *Race and Intelligence: Separating Science From Myth*, ed. Jefferson M. Fish (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002), 396.

<sup>40</sup>Barnett and Camilli, 1997, 19.

<sup>41</sup>Steven Barnett, National Institute for Early Education Research, e-mail message to Douglas Besharov, April 10, 2006.

<sup>42</sup>Barnett and Camilli, 1997, 19.

Hispanics, by nativity and English language in the home. It seems likely that the overall effects of Head Start would not have been statistically significant, but the UCLA team is silent on this.

The UCLA team discussed several possible reasons for Head Start's differential effects on the test scores and school performance of whites and African Americans. First, they hypothesized that the quality and content of Head Start programs varied, stating: "It is possible that African-American children are more likely to be served by inferior programs."<sup>43</sup> However, they were not able to test the hypothesis because they lacked data on individual programs.

Second, the impact of Head Start may depend on the child's home background and the quality of schools attended after Head Start graduation: "To the extent that African-American children come disproportionately from more disadvantaged homes, located in poorer communities, and attend more troubled schools, one might expect Head Start to have either smaller initial effects or effects that dissipate more quickly over time."<sup>44</sup> And yet, these are just the children for whom experimental evaluations seem to show early childhood programs having the greatest impact.

**Evaluator independence.** Currie, Thomas, and Garces are independent researchers in academia.

**Statistical significance/confidence intervals.** Statistical significance was measured and reported at the 1 percent, 5 percent, and 10 percent levels.

**Effect sizes.** Apparently effect sizes were either not calculated or not reported.

**Sustained effects.** The studies focused primarily on long-term effects.

**Benefit-cost analysis.** Although the UCLA team draws some conclusions about the benefits versus the cost of Head Start for different participants (see "Benefit-cost findings"), a formal benefit-cost analysis was apparently not performed.

**Cost-effectiveness analysis.** Apparently not performed.

---

<sup>43</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 356.

<sup>44</sup>Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 356.

## Commentary

**Editor's note:** Drs. Currie and Thomas provided us with preliminary comments. After making changes, we sent them the revised chapter to which they did not respond. The following are the preliminary comments they provided.

Douglas Besharov and his colleagues provide a comprehensive discussion of our research. Much of it is balanced and accurate. Some of the claims are false and several of their conclusions are flawed.

### Study Design

Perhaps the most important contribution of our work on Head Start using non-experimental data is the fact that the evidence regarding the effects of the program on child outcomes is remarkably consistent with results from experimental studies. Experimental and non-experimental studies are based on entirely different methodologies and there is no reason to expect they will come to the same basic answers by chance. The fact that they do in this particular case suggests that we can be more confident that the answer is probably right.

A key advantage of experimental studies is that subjects are randomly assigned to a treatment or control group. Non-experimental studies, on the other hand, make statistical assumptions in order to construct groups for which comparisons are likely to be informative. Our approach -- comparing siblings, one of whom attended Head Start and one of whom did not -- is one of many potential approaches. We think that it is particularly appropriate in this particular context because it enables us to compare two children who have a great deal in common but who have one key difference: one attended Head Start while the other did not. However, statistical assumptions in non experimental studies should be subjected to scientific scrutiny. For example, a key assumption in our work is that the sibling who attends Head Start would not be better off than the sibling who did not attend in the absence of the program. The validity of this assumption can be tested by examining characteristics of both siblings. Such a comparison suggests that children who attended Head Start are on average more disadvantaged than their siblings were at the same age.

One of our key results is that the effects of Head Start on cognitive indicators are the same for African American and white children at the time they exit the program. However, those benefits appear to fade out more quickly for African Americans, relative to whites. There are many plausible explanations for this observation. We have explored this issue in research that Besharov and his colleagues chose not to cite. We find, for example, that African American children who attended Head Start subsequently attend significantly lower quality schools than other African American children, while the same is not true for white Head Starters. This suggests -- indeed, only suggests -- that the fade out may have more to do with later life

experiences than Head Start. This suggestion is consistent with a large literature on child development.<sup>45</sup>

Besharov and his colleagues say:

Mark Lipsey and David Cordray of Vanderbilt University argued that the existing techniques to control for selection bias do not resolve the problem completely: “The major problem with the available statistical approaches to selection bias is the sensitivity of the results to the violation of the model assumptions, especially the requirements that all relevant variables be specified. Much work remains to be done on the question of which models are best for which circumstances.”<sup>46</sup> Thus, despite the careful analysis conducted by the Currie and Thomas, there is always some uncertainty regarding the findings of “quasi-experiments.”

There are two important points about this paragraph. First, there is uncertainty regarding findings in any study--experimental or non-experimental. Issues that can threaten the validity of an experimental design include: failures of randomization, differential attrition of treatments and controls, “Hawthorne effects”, and controls who seek treatment outside the experimental setting. The question is whether one believes the study design is informative. Second, Lipsey and Cordray are, in fact, arguing that reliance on evidence from experiment alone is fraught with difficulty. In fact, they make the case for compiling an evidence base drawing on both experimental and non-experimental sources. They argue that such an evidence base is the most likely to stand the test of time. That is precisely why we examined the effect of Head Start using non-experimental sources and it is why we conclude that taking the evidence from experimental and non-experimental studies in combination is overwhelming: Head Start yields benefits for some sub-groups of the population.

### **The Role of Statistics**

In many instance, the questions raised by Besharov and his colleagues indicate a poor understanding of scientific standards that are applied to statistical studies. For example, they claim that studies based on small samples cannot be considered reliable. Besharov and his colleagues conclude that our samples are (too) small and that “Moreover, such small numbers would also be a concern for experimental evaluations as well.” There are scientifically-founded statistical tests which measure whether a difference is the result of chance error or not. The smaller the sample, the more likely that such tests will not be able to distinguish measured effects

---

<sup>45</sup>Janet Currie and Duncan Thomas, “School Quality and the Longer-Term Effects of Head Start,” *Journal of Human Resources* 35, no. 4 (Fall 2000): 755–774.

<sup>46</sup>Mark W. Lipsey and David S. Cordray, “Evaluation Methods for Social Intervention,” *Annual Review of Psychology* 51 (2000): 349.

from chance error, and thus the more likely it is that effects will not be statistically significant. Thus the fact that we find significant results even in relatively small samples is evidence that the effects are well-determined and not due to chance error. In fact, many studies of early intervention programs to date have been based on small samples, and many of the effects that have been measured have not been significant from a statistical point of view although they are likely to be very important from the point of view of the future well-being of children at risk. In fact, a simple meta-analysis of the literature on Head Start would go some way to addressing this concern: the facts are that the overwhelming majority of estimated effects of Head Start on a plethora of child and young adult outcomes are positive. The probability that this is due to chance error is small.

### **What is the Interesting “Effect”?**

Besharov and his colleagues appear to think that the only interesting benefit is the benefit for the average child rather than the benefits for sub-groups. We find this to be a remarkably naive and poorly thought-out position. We can see no reason to think that Head Start should benefit everyone in the same way. Our research has shown large effects of Head Start on the language skills of Hispanic children, effects on the prevention of grade repetition among non-Hispanic Caucasian children, and reductions in the probability of ever having been booked or charged with a crime among African-Americans. The fact that Head Start does not appear to have the same effect on all children, is no reason for dismissing evidence of very real benefits to different subgroups. We believe that it is extremely important to understand the heterogeneity in the benefits associated with Head Start as it will point to the types of programs that are of greatest value for particular types of children and families. Armed with that evidence, we can begin to identify ways in which Head Start might be improved.

### **Conclusion**

Taking a step back and looking at the bigger picture, the evidence in support of the view that Head Start has benefits for some sub-groups of the population is overwhelming. It has been reported in experimental studies and in non-experimental studies. All of the evidence points in one direction: children at risk of losing ground benefit in multiple dimensions by participating in a quality preschool program like Head Start.

*Note:* This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to [welfareacademy@umd.edu](mailto:welfareacademy@umd.edu) or fill out the comment form at [http://www.welfareacademy.org/pubs/early\\_education/chapter6.html](http://www.welfareacademy.org/pubs/early_education/chapter6.html).