



SCHOOL of  
PUBLIC POLICY

# 13

## Head Start Summer Program (1965)

**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/chapter13.html](http://www.welfareacademy.org/pubs/early_education/chapter13.html).

# 13

## Head Start Summer Program (1965)

The Head Start program, started in 1965 as an eight-week summer program for three- and four-year-old children, is designed to help “break the cycle of poverty by providing preschool children from low-income families with a comprehensive program to meet their emotional, social, health, nutritional, and psychological needs.”<sup>1</sup>

Leon Eisenberg and C. Keith Conners, then of the Johns Hopkins School of Medicine (the “Johns Hopkins team”), evaluated the cognitive impact of the 1965 Head Start summer program in Baltimore, Maryland. One of many early evaluations of the Head Start program,<sup>2</sup> it is included here because at the time, it gained much attention among educators and in the media. This evaluation found that Head Start’s summer program produced immediate cognitive gains. There was no long-term follow-up, and the Johns Hopkins team speculated that the early gains might “fade out,” given the subsequent “educational impoverishment” characterizing inner-city schools. Although the evaluation was based on a comparison group design (rather than a randomized experiment), the evaluation is reviewed here because (1) it came so early in the life of Head Start, (2) the methodology is reasonably strong, and (3) the findings related to short-term impacts foreshadow those in subsequent research.

### Program Design

**Program group.** The program was targeted to children eligible for Head Start who were about to enter kindergarten. From its inception, the Head Start program has been targeted to children in families with incomes below the poverty line.

The Johns Hopkins team characterized the study population in Baltimore as “severely

---

<sup>1</sup>U.S. Department of Health and Human Services, Administration for Children and Families, Head Start Bureau, “Head Start History,” (Washington, DC: HHS, 2002), <http://www.acf.hhs.gov/programs/hsb/about/history.htm> (accessed November 8, 2005).

<sup>2</sup>See, for example, Westinghouse Learning Corporation and Ohio University, *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children’s Cognitive and Affective Development, Volume 1*, Report to the Office of Economic Opportunity (Athens, Ohio: Westinghouse Learning Corporation and Ohio University, 1969).

disadvantaged.” Nearly two-thirds of parents had no more than a tenth grade education; 30 percent of the families were on welfare; 38 percent of the mothers were single, widowed, divorced, or separated; and only 7 percent of children had previously been in day care.

**Services.** The Head Start summer program operated for six weeks. It was run by elementary school teachers with little training and experience working with preschoolers. The average class size was fifteen children and average daily attendance was over 90 percent.

**The Evaluation.** The Johns Hopkins team used a nonexperimental research design because “there were too few children in excess of available space to constitute an initial control group.”<sup>3</sup> Head Start children were tested upon program entry in June, again in August at program termination, and in September upon enrolling in kindergarten. The Johns Hopkins team selected the comparison group from children in the same kindergarten classes. They were tested only in September. Since the comparison group children lived in the same neighborhoods, the Johns Hopkins team assumed that they resembled the Head Start children on a range of socioeconomic characteristics. Cognitive impacts were measured using the Peabody Picture Vocabulary Test (PPVT) and the Draw-a-Person test (DAP).

### Major Findings

This evaluation found that Head Start’s summer program produced immediate cognitive gains.<sup>4</sup> There was no long-term follow-up, and the Johns Hopkins team speculated that the early gains might “fade out,” given the subsequent “educational impoverishment” characterizing inner-city schools.

**Cognitive.** The study was limited to investigating children’s cognitive gains. It compared the distribution of PPVT scores for the Head Start group in June and the comparison group children in September. The distributions were nearly identical, despite a ten-week age advantage for the comparison group children. The average PPVT score for the Head Start group in June was 32.63 compared to 33.65 for the comparison group in September. This led the Johns Hopkins team to conclude that, initially, the groups were not significantly different from each other in their knowledge of vocabulary.

The Johns Hopkins team found, however, that the Head Start children progressed steadily in subsequent test scores in August and September. They described the differences in September test scores between the Head Start and comparison group children as “striking.” The Head Start

---

<sup>3</sup>Leon Eisenberg and C. Keith Conners, “The Effect of Head Start on Developmental Processes,” in *Expanding Concepts in Mental Retardation: A Symposium*, ed. George A. Jervis (Springfield, IL: Charles C. Thomas, 1968), 117.

<sup>4</sup>Eisenberg and Conners, 1968, 116-122.

group's average score had risen to 39.74, more than 6 points higher than the comparison group's average score. These gains were especially large for the quartile of children who initially had the lowest scores. The size of the gain, which diminished as scores rose, was quite small for the highest quartile.

The findings using the DAP scores were very similar, again favoring the Head Start children. The average DAP score for the Head Start group in June was 7.71 compared with 8.91 for the comparison group in September. Despite the apparent initial advantage of the comparison group, by September, the Head Start group's DAP score had risen to 9.75 and exceeded the comparison group's score by nearly one point. This difference was also highly significant.

**School readiness/performance.** Data apparently either not collected or not reported.

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

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

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

**Crime/delinquency.** Data apparently either not collected or not reported.

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

**Economic outcomes.** Data apparently either not collected or not reported.

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

**Benefit-cost findings.** Apparently a benefit-cost analysis was not performed. At the time of the evaluation, Head Start cost approximately \$1,200 per child (in 2005 dollars).

### **Overall Assessment**

The evaluation was based on a comparison group design and is subject to selection bias. The findings related to short-term impacts, however, are credible and consistent with subsequent research.

**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. Within this context, the evaluation is appropriate.

**Program implementation.** No implementation problems were noted.

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

**Assessing statistical controls in experimental and nonexperimental evaluations.** The evaluation is based on a comparison group design, so selection bias is a potential problem. The Johns Hopkins team presented very little information about the socioeconomic characteristics of Head Start children and virtually none for the comparison group children. No statistical controls were used.

Both groups were drawn from the same neighborhood and the distribution of PPVT raw scores for Head Start children in June and comparison group children in September were comparable. Even if there were significant differences between the groups on a range of baseline characteristics, it seems unlikely that these differences would explain away the gains experienced by the Head Start group in the short, ten-week period between their enrollment and final test.

The two groups of children were less comparable with respect to their initial DAP scores. The Head Start group scored lower initially than did the comparison group (7.71 vs. 8.91). This would suggest that the Head Start children were more disadvantaged than the comparison group children, which might suggest the Head Start families were more disadvantaged. This would tend to understate Head Start's positive impact.

**Sample size.** The initial sample was 826 children for the PPVT and 920 children for the DAP test.<sup>5</sup> Thus, the sample was large enough to detect even relatively small differences in test scores.

**Attrition.** Since the comparison group was tested just once, attrition was not a factor. The number of program children taking the PPVT, however, declined 5 percent between June and September, from 425 to 403. The Johns Hopkins team compared each lost group to "a matched sample to make certain that selective loss did not account for the apparent gains; the lost groups did not differ from their matched controls on initial testing."<sup>6</sup> Thus, attrition does not appear to have been a serious problem.

**Data collection.** The data collection relied on two tests, the Peabody Picture Vocabulary

---

<sup>5</sup>The sample sizes reported in Tables 14-III and 14-IV of the evaluation report are somewhat inconsistent with the text and figures, raising some questions about the analysis. For example, the text (p. 118) indicates that with respect to the PPVT test, there were 712 children in both the Head Start (in June) and comparison (in September) groups. However, table 14-III indicates that there were 425 Head Start children (in June) and 402 comparison group children, for a total of 826 children. Figure 14-1 suggests yet a third sample size, with 425 Head Start children and 378 comparison group children, for a total of 803 children. These differences create some uncertainty about the validity of the comparisons since it is unclear how the findings differ depending on which sample members are included.

<sup>6</sup>Eisenberg and Connors, 1968, 120.

Test (PPVT) and the Draw-A-Person test (DAP), which are appropriate for the questions being studied. The tests were administered by trained nonprofessional volunteers, but the authors note that “satisfactory test-retest reliability has been demonstrated for the PPVT in a comparison of nonprofessional and professional examiners.”<sup>7</sup>

**Measurement issues.** Cognitive impacts were measured using the PPVT and the DAP. The Johns Hopkins team indicate that both have been shown to “correlate respectably with standard intelligence tests in middle-class populations.”<sup>8</sup> They contend that test repetition, which could have inflated the scores for the Head Start group, could not have explained the differences in program effects found.

**Generalizability.** The Baltimore Head Start summer program was not typical of today’s Head Start program. It was a short summer program during the first year the program operated, and the teachers running the program had little experience with preschoolers. Given the dramatic changes in the population eligible for Head Start and the changes in the program itself, the findings are clearly not generalizable to Head Start as it operates today. The study is noteworthy, however, because it was one of the first to establish that short-term cognitive gains result from Head Start participation and because of the Johns Hopkins team’s insight that such gains might “fade out.”

**Replication.** There have been numerous other evaluations of Head Start, including the summer program. The findings have been mixed, but many showed short-term cognitive gains that later “fade out.”

**Evaluator’s description of findings.** The Johns Hopkins team concludes that Head Start resulted in significant test gains for participating children. Unfortunately, they did not examine the longevity of these gains. For example, it is possible that the early kindergarten experience raised the abilities of the comparison group to the level of the Head Start children. To assess this would have required a subsequent test, sometime after September.

The Johns Hopkins team cautions that program gains might not last, “given the overcrowding, educational impoverishment, and generally negative attitudes toward the poor that characterize inner-city elementary schools.”<sup>9</sup> In other words, they speculate that the gains might “fade out,” but this possibility could not be assessed during the short follow-up period. Nevertheless, the findings of this evaluation are consistent with many other studies that show IQ and achievement test gains while children are participating in early childhood programs or shortly thereafter. Their prediction that early cognitive gains may not persist seems to have been confirmed

---

<sup>7</sup>Eisenberg and Connors, 1968, 117.

<sup>8</sup>Eisenberg and Connors, 1968, 117.

<sup>9</sup>Eisenberg and Connors, 1968, 122.

by many subsequent studies.

**Evaluator's independence.** Eisenberg and Conners were affiliated with the Johns Hopkins School of Medicine and were not directly involved in the administration of the Baltimore Head Start summer program.

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

**Effect sizes.** Apparently effect sizes were either not reported or calculated. The study reported significant differences in IQ between the control group and program group.

**Sustained effects.** The evaluation examined short-term post-intervention impacts.

**Benefit-cost analysis.** Apparently not performed.

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

## Commentary

**Editor's Note:** For each evaluation included in this report, we attempted to contact the senior evaluators to offer them the opportunity to respond to our assessment. When we contacted Drs. Leon Eisenberg and Keith C. Connors, they declined to provide comments.<sup>10</sup>

**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/chapter13.html](http://www.welfareacademy.org/pubs/early_education/chapter13.html).

---

<sup>10</sup>Dr. Keith Connors, e-mail message to Caeli Higney, January 19, 2006.