SCHOOL of PUBLIC POLICY

2
The Abecedarian Project

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

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 or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter2.html.

2
The Abecedarian Project

The Abecedarian Project, operating between 1972 and 1985,[1] was an intensive, center-based program that began working with children in infancy and continued through preschool (and the early school years for some children). It was designed to "test the hypothesis that providing socially disadvantaged children with an intellectually stimulating environment from early infancy could prevent the development of mild mental retardation."[2]

Craig Ramey and his colleagues at the Frank Porter Graham (FPG) Child Development Institute at the University of North Carolina, Chapel Hill (the "UNC team") conducted a random assignment evaluation of the program between 1972 and 1977, with follow-ups through age twenty-one. They concluded that the program successfully achieved positive and lasting gains on a wide range of cognitive and school-related outcomes. These gains, however, appear to have been concentrated among the subgroup of children whose mothers had IQs below 70 at the time of entry into the study, and some faded over time.[3] Moreover, these early gains did not lead to many improved outcomes in adulthood (when the children were age twenty-one), with, for example, no statistically significant increases in employment or reductions in criminal activity. Although the project was evaluated using random assignment, the post-random assignment refusal to participate in the evaluation of over 10 percent of families assigned to the program group raises the possibility of selection bias. In addition, the fact that the project was composed mainly of low-income, black children (at "high risk" of intellectual or academic failure) in an otherwise affluent area and the absence of successful replications raise questions about the

---

[1]In the last years of this period, the program was limited to school-related services provided to a subset of program and control group children.

[2]Craig T. Ramey and Frances A. Campbell, "Preventive Education for High-Risk Children: Cognitive Consequences of the Carolina Abecedarian Project," *American Journal of Mental Deficiency* 88, no. 5 (March 1984): 516. *See also,* The Carolina Abecedarian Project website, http://www.fpg.unc.edu/~abc (accessed June 28, 2010).

[3]Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 5, stating: "*The most vulnerable children benefitted the most from the preschool program.* We estimated the effects of the preschool treatment on intellectual development from 6 through 54 months of age (Martin, Ramey, & Ramey, 1990). The IQs of preschool treatment group children ranged from 8 to 20 points higher than those of control children when maternal mental retardation and home environment, both important factors in the occurrence of developmental delay, were statistically accounted for via multiple-regression analyses. The children who benefitted the most had mothers with IQs below 70."

generalizability of the findings.

Program Design

Program group. The Abecedarian Project recruited 122 children from 120 low-income families (including one set of twins and one pair of siblings) from local prenatal clinics and from families identified by the local Department of Social Services as falling within the eligibility criteria.

Eligibility was based on a thirteen-factor High Risk Index, which included mother's and father's educational level, family income, father's absence, welfare receipt, and other factors related to children's lower levels of intellectual functioning and/or academic failure. All families whose scores exceeded a threshold level for risk and who agreed to participate were included in the experiment. Infants who were subsequently observed to have an obvious neurologic disorder, however, were dropped from the study.

At the time of enrollment, the mean age of the mothers in the program group compared to the control group were sightly younger on average (about 19.7 years vs. about 20.4 years), had about the same average educational level (about 10.6 years vs. 10.1 years), were more likely to be single mothers (83 percent vs. 75 percent), less likely to be black (94 percent vs 100 percent), and more likely to be first-time mothers (67 percent vs. 60 percent). All differences were not statistically significant.[4]

Services. The Abecedarian Project provided a full-day (six to eight hours per day), full-year preschool program, beginning with children as young as six weeks of age (with an average age of entry of 4.4 months) and continuing until the children entered kindergarten.[5] Staff-to-child ratios were 1:3 for infants and children up to age two and 1:6 for two- to five-year-olds. Highly trained and well supervised staff taught age-appropriate curricula that emphasized the development of communication skills. An infant curriculum for children up to age three focused on language, motor, social, and cognitive skills. For the older children, staff developed individualized education programs that emphasized language, cognitive, social, and physical development. Parents were offered varying amounts of social services including personal counseling, and they were encouraged to participate in group sessions on topics such as parenting and family development.

---

[4]Craig T. Ramey and Frances A. Campbell, "Preventive Education for High-Risk Children: Cognitive Consequences of the Carolina Abecedarian Project," *American Journal of Mental Deficiency* 88, no. 5 (March 1984): 515–523.

[5]The Abecedarian Project offered as much as 10,000 hours of preschool intervention services per child (eight hours per day, five days per week, fifty weeks per year, for five years) compared to the 720 hours received in a typical Head Start program at that time (four hours per day, five days per week, for thirty-six weeks).

To address the possibility of "fade out"—or loss of early gains, a pattern observed in other experiments—a second phase of the program was created for about half of the children from kindergarten through second grade (K-2). (Half of the program group and half of the control group were randomly assigned to this component.) This second phase paired families with experienced home/school resource teachers (HSTs) who visited the family every two weeks, providing additional educational material, designing educational activities for parents to conduct with their children, and offering parental support and advocacy. In addition, the HSTs visited the child's classroom every two weeks to ensure that the supplemental materials provided at home complemented the child's school curriculum and emphasized areas in which the child needed extra help. During some summers, participants in the K–2 program group also attended summer camp as a further developmental supplement.[6]

The Evaluation. The UNC team conducted the evaluation. Between 1972 and 1977, they randomly assigned 120 families (with 122 children) to either a program group or to a control group. Follow-up was conducted periodically until participants turned twenty-one. When the children reached age five, ninety-six remained in the sample and were randomly assigned again to one of four groups: (1) a preschool plus school-age program group; (2) a preschool-only program group; (3) a control group; and (4) a preschool control plus school-age program group. This allowed the UNC team to compare child outcomes for: early intervention, early intervention plus transitional school support, school-age intervention only, and no intervention. Leonard Masse and Steven Barnett, both at the National Institute for Early Education Research at Rutgers University, conducted a benefit-cost analysis published in 2002.[7]

Major Findings

The Abecedarian evaluation findings suggest that the program successfully achieved positive and lasting gains on a wide range of cognitive and school-related outcomes, including IQ, reading, and mathematics achievement scores.[8] However, these gains became ambiguous as

[6]Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 2–14, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June 28, 2010).

[7]Leonard N. Masse and W. Steven Barnett, *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention* (New Brunswick, NJ: National Institute for Early Education Research, 2002), http://nieer.org/resources/research/AbecedarianStudy.pdf (accessed June 28, 2010).

[8]Unless otherwise noted, all findings are from: Frances A. Campbell, Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson, "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project," *Applied Developmental Science* 6, no. 1 (January 2002): 42-57; Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 2–14, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June

time went on, as described below. At age fifteen, the Abecedarian group had less grade retention and fewer special education placements. These early gains, however, did not lead to many improved outcomes in adulthood (when the children were age twenty-one), with, for example, no statistically significant differences in high school graduation rates, employment, or criminal activity.

Cognitive. The evaluation included findings for both IQ and achievement scores.

IQ. Differences in cognitive development between the program and control groups became apparent almost immediately and increased throughout the first three years of the preschool treatment period (see table 1). For example, no statistically significant difference in IQ was observed when the children were three months old. By six months, however, the program group averaged IQs that were 5 points higher than the control group and, by age three, their IQs were 17 points higher. After age three, these differences diminished, and at age five, the IQ gains were only 7 points. At age twelve, the program group's average IQ was 6 points higher than the control group, but there were no statistically significant differences at age eight or age fifteen. At age twenty-one, however, the program group showed gains of about 5 points. (These final gains were limited to female participants, who gained an average of 8 points [90 vs. 82].)

The UNC team points out that some subgroups gained more than others.[9] In particular, the children who had mothers with IQs below 70 had the largest gains.[10] Among this subgroup, at age fifteen, the effect of preschool on child IQ was about 10 points (91 vs. 81).[11] Given the small sample size (six program individuals and six control individuals), this finding should be viewed as suggestive only. However, it is consistent with two themes underlying many studies: (1) Only those children who are more "at risk" seem to show the most substantial benefits from early childhood programs, and (2) first-time mothers seem to be the most amenable to program services and interventions.

28, 2010); and Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 231–244.

[9]Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 2–14, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June 28, 2010).

[10]As will be discussed later, this adds support to the theory behind the Milwaukee Project that lower IQs among at-risk children are often a function of being raised by "an intellectually limited caregiver."

[11]Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 10, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June 28, 2010).

Achievement. The reading and mathematics achievement scores were consistently higher for the program group throughout school to age twenty-one (see table 2). At age twenty-one, the program group outscored the control group by about 5 points on the Woodcock-Johnson Broad Reading (93 vs. 88) and Broad Mathematics (89 vs. 84) tests. These differences translate into a grade-equivalent increase in reading scores of two years (11.1 vs. 9.2) and in math scores of more than one year (9.2 vs. 7.9).

Table 1. Abecedarian Project: IQ Effects

Age	Program group	Control group	Difference (percentage points)
3 months	95	95	—
6 months	107	101	6
9 months	110	110	—
12 months	111	105	6
18 months	108	90	18
2 years	96	85	11
3 years	101	84	17
4 years	102	89	13
5 years	101	94	7
6.5 years	98	93	5
8 years	98	94	—
12 years	95	89	6
15 years	96	90	—
21 years	90	85	5

Sources: Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 2–14, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June 28, 2010); Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 231–244.

Notes: Only statistically significant differences are reported. "—" indicates that the difference is not statistically significant at the 5 percent level. The Bayley Mental Development Index (MDI) was used for measurements between three and eighteen months; the Stanford-Binet was used for measurements at ages two to four; the Wechsler Preschool and Primary Scale of Intelligence was used at age five; the Wechsler Intelligence Scale for Children-Revised was used at ages six and one-half, eight, twelve, and fifteen; and the Wechsler Adult Intelligence Scale-Revised was used at age twenty-one.

Table 2. Abecedarian Project: Effects on Achievement Test Scores

	Reading scores			Math scores		
Age (years)	Program group	Control group	Difference (percentage points)	Program group	Control group	Difference (percentage points)
8	94	85	9	98	92	6
12	91	85	6	92	87	5
15	94	89	5	94	87	7
21	93	88	6	89	84	5

Source: Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 234.

Notes: Only significant differences are reported. "—" indicates that the difference is not statistically significant at the 5 percent level. The Woodcock-Johnson Achievement Test for Reading and Math was used.

There were no statistically significant differences by gender or between the preschool-only and preschool plus K-2 program groups. As with IQ, the effects were largest for the subgroup of children with mothers with low IQs.

School readiness/performance. At age fifteen, the Abecedarian program group had less grade retention (31 percent vs. 54 percent) and fewer special education placements (25 percent vs. 48 percent). At age twenty-one, however, there was no statistically significant difference in high school graduation rates between the program and control group. Frances Campbell, a senior scientist at the FPG Child Development Institute, explains that the absence of a statistically significant impact here may be due to the fact that the children lived in an area where it was unusual for even high-risk students to drop out.[12] The program group, however, experienced a 21 percentage point increase in the percent who had ever enrolled in a four-year college (36 percent vs. 14 percent).

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

Health. At age twenty-one, there were no statistically significant differences in various health-related behaviors, including being a regular smoker, drinking any alcohol in the past thirty days, or binge drinking in the past month.

[12]Frances Campbell, Frank Porter Graham Child Development Institute, e-mail message to Peter Germanis, January 12, 2001.

Behavior. After three years of school, there were no statistically significant differences in children's behavior, based on parental responses to the Child Behavior Checklist.

Crime/delinquency. At age twenty-one, there were no statistically significant differences in self-reported misdemeanor convictions, felony convictions, or incarceration. While the Abecedarian group was less likely to have smoked marijuana in the past month (18 percent vs. 39 percent), there was no statistically significant difference in the use of cocaine or other drugs.

Besides the possibility or probability that the program had no impact on crime, one reason for the absence of any apparent impact on this measure may have been the relatively low crime rate in the area in which the project operated, so that the rates were low for both the control and program groups. Another possibility is that the project did not provide services to address this issue. Stevens Clarke, professor of Public Law and Government at the University of North Carolina at Chapel Hill, examined the Abecedarian data. He explains that one difference between the Abecedarian Project and other programs that have found impacts in this area (most notably, the High/Scope Perry Preschool Project and the Syracuse Family Development Program) is that the latter programs had a more intensive family involvement component during the treatment phase.[13]

Early/nonmarital births. The program group was more likely to delay having a first child. Only 26 percent of program group children reported becoming parents in their teens, compared to 45 percent of control group children. Among those who did have a child by age twenty-one, the mean age at the birth of the first child was 19.1 years for the program group, compared to 17.7 years for the control group.

Economic outcomes. At age twenty-one, the employment rate was higher for the program group (64 percent vs. 50 percent), but the difference was not statistically significant. There were, however, differences in the reported skill levels of employment, with 47 percent of the program group reporting that they were in a skilled job, compared to 27 percent of the control group. (Skill level was measured by the Hollingshead index.) There were no statistically significant differences in various measures of self-sufficiency, such as maintaining a home, providing full support for themselves, or owning a car.

Effects on parents. Mothers who were teens when they enrolled in the project were much more likely to complete high school than were teen mothers in the control group (46 percent vs. 13 percent). In addition, they were more likely to be employed when the children were fifteen years old (92 percent vs. 66 percent). There was little difference for mothers who were not teens at the time of enrollment.

[13]Frances Campbell, Frank Porter Graham Child Development Institute, e-mail message to Peter Germanis, January 12, 2001.

Benefit-cost findings. Early childhood intervention programs are often justified on the assertion that they produce savings that exceed their costs. Masse and Barnett conducted a preliminary benefit-cost analysis of the Abecedarian project.[14] Their analysis included benefits and costs to participants, their families, government, and the program itself. In this paper, we only consider the project from the government's perspective, to determine whether a program like the Abecedarian project could save more in government expenditures than it cost, often called the "taxpayer" perspective.[15]

Since program costs are incurred "up front" while some benefits and costs appear only later, the rate at which society is willing to tradeoff future benefits and costs for current benefits and costs (the discount rate) affects the estimated "present value" of benefits and costs. For the purpose of this analysis, all estimates are adjusted to 2005 dollars, using a 3 percent annual discount rate.

Costs. Masse and Barnett estimated the total, five-year cost of the Abecedarian Project to be about $73,000 per child (undiscounted, in 2005 dollars).[16] In a benefit-cost analysis, however, the focus is not on total costs, but on a comparison of *net benefits* to *net costs*. In an experimental evaluation, for example, this means comparing the difference in dollars between program and control group members for both benefits and costs. The net cost is the additional cost beyond the cost of child care for children in the control group. Masse and Barnett based their estimate of the cost of child care services for the control group children on: (1) participation data documenting their use of center-based child care; (2) estimates of the hours spent in center-based and informal care arrangements (based on national data from the National Household Education Survey); (3) the price of child care (during preschool years only) for relative and non-relative care (estimated using data from the National Child Care Survey) and for center-based care (estimated at the average rate reported for centers accredited by the National Academy of Early Childhood Programs in a U.S. Government Accountability Office report). In addition to the estimated out-of-pocket child care expenditures, Masse and Barnett added a cost for "parental care" for parents

[14]W. Steven Barnett and Leonard N. Masse, "Comparative Benefit-Cost Analysis of the Abecedarian Program and Its Policy Implications," *Economics of Education Review* 26 (2007): 113–125, http://nieer.org/resources/research/BenefitCostAbecedarian.pdf (accessed October 20, 2010).

[15]Other perspectives might also be examined. The "participant" perspective compares the benefits of a program to participants relative to the program costs they bear. The "social" perspective takes the point of view of the society as a whole, which reflects both the "participant" and "taxpayer" perspectives. For example, the "social" perspective would count increased participant earnings as a social benefit and increased administrative costs as a social cost. In some cases, the benefits and costs counted by the "participant" and "taxpayer" perspectives offset each other. For example, a program that reduces welfare payments would produce a cost to participants, but an offsetting gain to taxpayers.

[16]Leonard N. Masse and W. Steven Barnett, *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention* (New Brunswick, NJ: National Institute for Early Education Research, 2002), 45, http://nieer.org/resources/research/AbecedarianStudy.pdf (accessed June 28, 2010).

who stay at home with their children.[17] As a result, they estimated the five-year, net cost of the program (in 2005 dollars, using a 3 percent discount rate) to be about $38,900 per child.

Benefits. Estimated benefits were calculated in four main categories: earnings and fringe benefits of the participants' mothers, the participants themselves, and future generations; education-related costs; health; and welfare use.

Masse and Barnett estimated the effect of the project on the child participants' lifetime earnings. Lacking direct information about earnings, they estimated earnings impacts based on the differences in educational attainment at age twenty-one between those in the program group and those in the control group. They also estimated future educational attainment and then estimated earnings effects through age sixty-five. Using this approach, they estimated lifetime earnings gains of about $40,700 per child participant. In addition, they estimated the impacts on earnings for the mothers (while they were between the ages of twenty-six and sixty) based on differences in earnings when the child participants were twelve, fifteen, and twenty-one years of age. They assumed no change in the earnings differential between the two groups at younger and older ages and thereby estimated lifetime compensation gains of $74,600 per mother. Finally, Masse and Barnett estimated the program's effect on the earnings of future generations to be about $6,200 per participant. The total lifetime compensation gains for children, their mothers, and future generations were estimated at about $121,500. This very large impact, however, was achieved only by relying on highly uncertain assumptions about educational attainment and earnings differentials. Moreover, from the government's budgetary standpoint, these earnings gains are relevant only in so far as they affect tax revenues and public outlays on welfare and other social programs. The impact of the project on these outcomes was largely ignored, however.

The savings (and costs) associated with schooling, including special education, were based on school attendance patterns, using national estimates of the costs of regular education and special education. The savings from reduced grade retention and special education placement were estimated to be about $9,600 per child participant. However, the added costs of higher education were estimated to be about $8,800 greater, since program group participants were more likely to be attending postsecondary educational institutions at age twenty-one. The net effect on education-related costs, thus, was estimated to be a net savings of about $800 per child participant.

At age twenty-one, the program group was less likely to use welfare (8 percent vs. 16 percent), although this difference was not statistically significant. Nevertheless, Masse and Barnett estimated savings in welfare administrative costs of $215 per participant. (The authors did not estimate savings from welfare benefits as "a reduction in welfare payments to program participants represents a transfer of money to the general taxpayer and does not change total

[17] An hour of parental care was valued at the price of non-relative care.

social benefits associated with the program.")[18]

The final component of the benefit-cost analysis was based on a relatively large, but not statistically significant, program/control difference in those reporting to be "regular smokers" as young adults (39 percent vs. 55 percent). Masse and Barnett used estimates of life expectancy for those who were or had been regular smokers at age twenty and the value of a life.[19] They estimated savings of $19,300 for each child. The impact of smoking on total government costs, however, was not estimated. While publicly funded medical costs may be higher for smokers, transfer payments such as social security may be lower due to their shorter life expectancy. Also, reductions in smoking have been associated with improved birth outcomes, which could lead to reductions in future health and special education costs for the next generation. Insufficient detail is provided to assess this aspect of their estimates.

Notably, their benefit-cost analysis did not include savings due to reductions in crime-related costs. In the High/Scope Perry Preschool Project (see chapter 14), the largest societal savings came from reductions in crime.[20] As noted above, however, there were no statistically significant crime effects in the Abecedarian Project, so crime-associated public costs were estimated to have zero dollar value.

Benefit-cost ratio. Masse and Barnett estimated that the Abecedarian Project had a net present value of $102,915 per child, saving $3.64 for every $1.00 spent. Table 3 summarizes the benefits and costs of the program. Indeed, a press release from the FPG Child Development Institute asserts that "For every dollar spent on high-quality early education programs, taxpayers can expect four dollars in benefits."[21]

Most of the benefits in their analysis, however, accrued to individuals in the form of greater earnings, rather than to the taxpayer. (Moreover, most of the earnings gain was projected and had not yet materialized.) Using their data, however, it is possible to estimate roughly the benefits and costs to the government. The Masse and Barnett estimate of the participants' lifetime compensation gains can be used to estimate increased tax revenues. For example, in the High/Scope Perry Preschool evaluation, Barnett assumed that increased tax revenues would be

---

[18]W. Steven Barnett and Leonard N. Masse, "Comparative Benefit-Cost Analysis of the Abecedarian Program and Its Policy Implications," *Economics of Education Review* 26 (2007): 119, http://nieer.org/resources/research/BenefitCostAbecedarian.pdf (accessed October 20, 2010).

[19]The question was worded: "How old were you when you started smoking on a regular basis?"

[20]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).

[21]Frank Porter Child Development Institute, University of North Carolina, Chapel Hill, "High-Quality Child Care Returns Far More Than Cost - New Report,"(press release, Frank Porter Child Development Institute, November 20, 2002).

equal to 25 percent of the compensation gain.[22] Applying the same assumption here results in estimated added tax revenues of about $10,200 from participants, about $1,550 from their children, and $18,650 from the mothers of the child participants. In addition, as noted earlier, the estimate of the program's net cost was based in part on an imputed parental cost for the control group, which did not involve governmental payments. Excluding this "cost" raises the net cost of the Abecedarian intervention to $49,141. (Even this is a conservative estimate of the government's net costs, since it assumes that the government absorbed all of the child care costs for control group children in center-based child care and other arrangements, which would be unlikely. Those payments made by parents themselves should be excluded from a benefit-cost analysis from the taxpayer perspective.) After making these modifications, this approach suggests that the Abecedarian Project would save taxpayers only about 63 cents for each dollar spent. Given the uncertainty of the various estimated benefits of the program, this casts a large question mark over Abecedarian's potential impact on government spending.

[22]W. Steven Barnett, "Cost-Benefit Analysis," in *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27* (Ypsilanti, MI: High/Scope Press, 1993), 143–73.

Table 3. Abecedarian Project: Estimated Net Benefits

	Net benefits to:	
	Society (Masse/Barnett)	**Taxpayers** (authors' calculations)
Benefits		
Education		
K-12	$9,600	$9,600
Higher education	-$8,800	-$8,800
Taxes		
Participants	—	$10,175
Children of child participants	—	1,550
Maternal	—	18,650
Earnings		
Participants	$40,700	—
Children of child participants	$6,200	—
Maternal	$74,600	—
Smoking	$19,300	?
AFDC	$215	$215
Total benefits	$141,815	$31,175
Program cost	$38,900	$49,141 (less parental cost)[a]
Net present value	$102,915	-$17,976
Benefit-cost ratio	$3.64/1	$0.63/1

Source: Adapted from Leonard N. Masse and W. Steven Barnett, *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention* (New Brunswick, NJ: National Institute for Early Education Research, 2002), http://nieer.org/resources/research/AbecedarianStudy.pdf (accessed June 28, 2010).
Notes: In 2005 dollars discounted at 3 percent.
[a] The cost estimate used by Masse and Barnett included an imputed "parental cost" to families in the control group. This does not involve governmental payments. Excluding this "cost," raises the net cost to $49,141 (authors' calculation).

Of course, this is still a crude estimate, as it is based on many assumptions and leaves out many potential benefits and costs. Nevertheless, it highlights the sensitivity of these findings to the assumptions behind them—especially concerning what is and is not included in the benefit-cost calculation—and the resulting uncertainty of benefit-cost claims.

Abecedarian's relatively small sample size also adds uncertainty. In a reanalysis of the High/Scope Perry Preschool Project and the Parental/Early Infancy Project, Lynn Karoly and her colleagues at RAND estimated savings in both projects of about $31,485 (in 2005 dollars), but

cautioned that the true savings had a two-thirds chance of being between $24,000 and $36,000 (and $18,000 to $42,000 using the standard 95 percent confidence interval).[23] Since the Abecedarian Project's sample size is in the same range as these other studies, similar confidence bands would probably apply here as well. Such wide confidence intervals also suggest that considerable caution should be used in making claims about benefit-cost findings.

Hence, the benefit-cost findings are too speculative and uncertain to provide definitive evidence on whether the program pays for itself. And, in fact, the principal investigators have not made claims based on them.

Overall Assessment

The project was evaluated using random assignment and the researchers involved described the evaluation procedures in considerable detail. However, the post-random assignment refusal to participate in the evaluation of over 10 percent of families assigned to the program group raises the possibility of selection bias. In addition, the fact that the project was composed mainly of low-income, black children (at "high risk" of intellectual or academic failure) in an otherwise affluent area, and the absence of successful replications raise questions about the generalizability of the findings.

Program theory. According to the UNC team:

> The Abecedarian Project was theoretically grounded with a conceptual framework based on General Systems Theory. From such a perspective, child development would be viewed as an ongoing process of interactions among hierarchical systems, ranging from that of the individual and factors that directly affect physical survival, to the psychological, involving interactions with caregivers, social systems in homes, schools, and neighborhoods, and societal forces. . . . [The Theory] provides a framework showing how changing the early environment, through supporting positive changes in children, could have long-term effects on later accomplishments.[24]

Based on this conceptual framework, one of the original purposes of the Abecedarian

[23]Authors' estimate based on Lynn A. Karoly, Peter W. Greenwood, Susan S. Everingham, Jill Hoube, M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders, and James Chiesa, *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions* (Santa Monica, CA: RAND, 1998).

[24]Frances A. Campbell, Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson, "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project," *Applied Developmental Science* 6, no. 1 (March 2002): 43.

project was to "demonstrate that sociocultural retardation can be prevented."[25] According to Herman Spitz, former director of the Research Department at the E.R. Johnstone Training and Research Center in Bordentown, New Jersey, "As the Project director and his colleagues saw it, this type of mental retardation was a 'sociocultural' disorder, meaning that most children who were mentally retarded and from poor environments were not born retarded. Rather, their intelligence gradually descended into the mentally retarded range because of the inadequate and confused inputs they experienced as they matured."[26]

But, as Spitz points out, in later publications, the purpose was broadened to include investigating "the degree to which mild retardation and school failure could be prevented."[27] Afterward, the purpose appears to have shifted to preventing "suboptimal cognitive development" and "school failure." Spitz argues that this shift was because the Abecedarian team "could not proved that they had prevented mental retardation unless the control group dropped into the mentally retarded range. The failure of the mean score of the control group to drop to the mentally retarded level renders problematic any strong claim that the Abecedarian Project prevented mental retardation."[28]

The UNC team carefully examined a variety of outcomes, including IQ, achievement, and school readiness. Thus, despite the program's evolving purpose, the evaluation seems appropriate within the context of the underlying theory.

Program implementation. Relatively little information has been published regarding implementation, for example, the participation patterns of program participants (and control group children) in various program-related services. In any event, apparently, no serious implementation problems have been reported.

Assessing the randomization. Four cohorts of infants born between 1972 and 1977 were randomly assigned in the study. The UNC team reports that: "Assignment was accomplished by

---

[25]Craig T. Ramey, M.C. Holmberg, J.H. Sparling, and A.M. Collier, "An Introduction to the Abecedarian Project," in *Infant Education: A Guide for Helping Handicapped Children in the First Three-Years*, ed. Bettye M. Caldwell and Donald J. Stedman (New York: Walker, 1977), as quoted in Herman H. Spitz, "Does the Carolina Abecedarian Early Intervention Project Prevent Sociocultural Mental Retardation?" *Intelligence* 16, no. 2 (April-May 1992): 17.

[26]Herman S. Spitz, "Attempts to Raise Intelligence," in *The Development of Intelligence*, ed. Mike Anderson (East Sussex: Psychology Press, 1999), 283.

[27]Frances A. Campbell and Craig T. Ramey, "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families, *Child Development* 65, no. 2 (April 1994): 685.

[28]Herman S. Spitz, "Attempts to Raise Intelligence," in *The Development of Intelligence*, ed. Mike Anderson (East Sussex: Psychology Press, 1999), 282-283.

pair-matching children on the high-risk index scores and from a table of random numbers assigning one number of each pair to the experimental condition."[29] Most articles published by the UNC team describe the sample as having 111 children, 57 randomly assigned to the preschool program group and 54 to preschool control group. In fact, 122 children were originally assigned to the research sample, but 11 dropped out before participation began.[30] If such attrition were random, this would not have been a serious problem, but this does not appear to have been the case. Of the eight mothers that refused to participate, seven were in the program group and one was in the control group. According to the UNC team, "The higher rate of rejection by families offered the preschool treatment was generally related to mothers wanting to care for infants in the home."[31] Simply dropping these cases undermined the integrity of random assignment.[32] As Robert St.Pierre, former vice president and principal associate at Abt Associates Inc., notes:

> If the cases had been kept in the data collection then the researchers could have done the analysis with and without them, providing empirical evidence as to the importance of these cases. At the very least, it would have been helpful to know how the cases that were dropped compare to the cases that were retained. I assume that there are important differences, as indicated by the mothers' interest in caring for their children at home.[33]

Three other children were dropped. One retarded infant was removed from the program group because the project was not intended to serve retarded children. Two child abuse cases were removed from the control group so they could receive the project's services, but they were

[29]Craig T. Ramey and Frances A. Campbell, "The Carolina Abecedarian Project: An Educational Experiment Concerning Human Malleability," in *The Malleability of Children*, ed. James J. Gallagher and Craig T. Ramey (Baltimore, MD: Paul H. Brookes, 1987), 127–139.

[30]Initially, 122 families were contacted and 121 agreed to participated. One mother miscarried, so 120 families were randomly assigned. The sample included one set of identical twins and one pair of siblings, raising the number of children to 122.

[31]Frances A. Campbell and Craig T. Ramey, "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention," *American Educational Research Journal* 32, no. 4 (Winter 1995): 748.

[32]The families should have been kept in the sample, even if it meant diluting the intervention. This way, if there had been an impact, we would have known that it was real. Even though the seven mothers in the program group did not want to send their children to day care, they might have been willing to participate in the follow-up data collection. Indeed, the incentives given control group members (free infant formula and disposable diapers) could have been offered them as well to maintain their cooperation.

[33]Robert St.Pierre, e-mail message to Peter Germanis, March 19, 2001.

removed from the research sample.[34] In sum, eight out of the eleven removed cases were in the program group, creating an immediate differential attrition rate. It would have been important to know more about the characteristics of these families to see whether this introduced any bias. Although the UNC team reports that they found no systematic differences between refusing parents and other parents, it appears that they did not have very much data on which to make this comparison. As suggested by the following discussion, it is misleading to exclude these cases in discussions about the comparability of the research groups, attrition, and potential biases, but it is too late to modify the analytic work to reflect their outcomes.

There are several other issues related to the random assignment of children. First, two siblings were automatically assigned to the program group to avoid a family having children in different groups. The assignment of these siblings is somewhat problematic, since they have no randomly assigned counterpart, but given their small number, it is unlikely that this created a serious bias.[35]

Further, as noted earlier, most reports indicate that 111 children were randomly assigned, with 57 assigned to the program group and 54 assigned to the control group. Adding back the dropped cases (discussed earlier) resulted in eight more cases in the program group (or sixty-five children) and three more in the control group (or fifty-seven children). The two siblings in the program group were automatically added to this group, suggesting that sixty-three families were assigned to the program group and fifty-seven to the control group. Yet, because children were paired and then randomly assigned, there should have been sixty families in each group; the fact that there were not suggests something else may have been amiss that was not reported.

Table 4 shows the baseline characteristics of 109 families (representing 111 children). (The Abecedarian Project evaluations did not present baseline characteristics on all 122 children.) Although the two groups look roughly similar, and there were no statistically significant differences on these baseline measures, the program group mothers were about nine months younger, yet had half-a-grade more education. Unfortunately, most reports about the study present data on relatively few baseline characteristics, despite the fact many others were apparently collected and used to calculate the high-risk index.

[34]Arguably, the child abuse cases too should have been retained in the research sample as control cases, even if it meant that some contamination crept in, since removing them undermined random assignment. While it is unfortunate that these cases were removed, it is notable that it was only two cases.

[35]This problem could be avoided simply by limiting the analysis to the older sibling.

Table 4. Abecedarian Project: Baseline Characteristics of Participating Mothers

Characteristic	Program Group	Control Group
Maternal age (years)	19.56	20.28
Maternal education (years)	10.45	10.00
Maternal IQ	85.49	84.18
Percent black	96%	100%
Sample size	55	54

Source: Frances A. Campbell and Craig T. Ramey, "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention," *American Educational Research Journal* 32, no. 4 (Winter 1995): 749.

In addition, the pattern of differences varied considerably across the cohorts that were randomly assigned. Although this information has generally not been presented separately for each cohort, the UNC team reported information on the first two cohorts from the Abecedarian study, using data from fifty-four randomly assigned children. In that study, they report the following baseline characteristics: "Maternal ages ranged from 14 to 44; the mean was 18.5 for mothers of treated children and 21.5 for mothers of controls. The educational levels of mothers (in grades completed) ranged from 6 to 12 with the mean being 10 years in both groups. Average maternal IQs were in the low 80s in both groups."[36]

The age difference between the mothers in the program group and the control group was three years for the early cohort, but narrowed to less than a year by the time the sample was completely enrolled (19.56 vs. 20.28 years). The mean age for the later cohorts, therefore, must have been approximately 20.7 years for the Abecedarian group and 19.2 years for the control group. In other words, for the first two cohorts, the mothers in the Abecedarian group were three years younger and for the next two cohorts, they were about one and one-half years older.[37] It is not clear whether this change in the age of mothers for the program group was due to random chance or removal of some cases, as described earlier.

Some researchers have speculated that preexisting differences between program and

[36]Craig T. Ramey, Bruce Dorval, and Lynne Baker-Ward, "Group Day Care and Socially Disadvantaged Families: Effects on the Child and the Family," in *Advances in Early Education and Day Care*, vol. 3, ed. Sally Kilmer (Greenwich, CT: JAI Press, 1983), 89.

[37]Spitz observes that the four cohorts allowed "the directors [of the evaluation] to compare cohorts, which they started to do. But when the results for cohorts turned out differently from each others, the project directors stopped comparing them, suddenly deciding the cohorts were too small. They were indeed small, but they were each getting the same curriculum, so they should have replicated each other. They didn't." Herman Spitz, e-mail message to Peter Germanis, March 19, 2002.

control group children account for some of the observed impacts. For example, Spitz has noted that the IQ gain at six months was only slightly lower than at age five, leading him to question the value of the additional "4.5 years of continuing intervention, 5 full days a week, 50 weeks a year."[38] He further wondered whether the differences at six months were due to the intervention or to differences in preexisting characteristics of the children and their families.[39] Richard Herrenstein, then professor of psychology at Harvard University, and Charles Murray, Bradley Fellow at the American Enterprise Institute, offer a similar explanation for this anomaly:

> Perhaps the intervention had achieved all its effects in the first months or the first year of the project (which, if true, would have important policy implications). Or perhaps the experimental and control groups were different to begin with (the sample sizes for any of the experimental or control groups was no larger than fifteen and as small as nine, so random selection with such small numbers gives no guarantee that the experimental and control groups will be equivalent). To make things still more uncertain, test scores for children younger than 3 years are poor predictors of later intelligence test scores, and test results for infants at the age of 3 or 6 months are extremely unreliable. It would therefore be difficult in any case to assess the random placement from early test scores. The debate over the results is ongoing . . . .[40]

Assessing statistical controls in experimental and nonexperimental evaluations. Given the use of random assignment, concerns about selection bias in the Abecedarian Project are relatively small. As previously noted, the program and control groups were comparable on a number of maternal characteristics, including high-risk index scores, maternal age, maternal education, maternal IQ, race, and percent living in an intact family. But, they were not identical, and the UNC team did not control for these differences, as has been common practice in many large-scale evaluations.[41] In addition, as noted, the exclusion of eleven children whose mothers

---

[38]Herman H. Spitz, "When Prophecy Fails: On Ramey's Response to Spitz's Critique of the Abecedarian Project," *Intelligence* 17, no. 1 (January–March 1993): 20.

[39]For more on the debate regarding this issue, *see* Herman H. Spitz, "Does the Carolina Abecedarian Early Intervention Project Prevent Sociocultural Mental Retardation?," *Intelligence* 16, no. 2 (April–June 1992): 225–237; Craig T. Ramey, "A Rejoinder to Spitz's Critique of the Abecedarian Project," *Intelligence* 17, no. 1 (January–March 1993): 25–30; Herman H. Spitz, "When Prophecy Fails: On Ramey's Response to Spitz's Critique of the Abecedarian Project," *Intelligence* 17, no. 1 (January–March 1993): 17–23; and Herman H. Spitz, "Spitz's Reply to Ramey's Response to Spitz's First Reply to Ramey's First Response to Spitz's Critique of the Abecedarian Project," *Intelligence* 17, no. 1 (January–March 1993): 31–35.

[40]Richard Herrenstein and Charles Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York: The Free Press, 1994), 407.

[41]Examples include the evaluations of the Infant Health and Development Program (see chapter 16), the Comprehensive Child and Development Program (see chapter 3), Early Head Start (see chapter 6), and the Head Start Impact Study (chapter 13).

refused to participate after being randomly assigned presents a source of potential bias.

Sample size. The original sample in the Abecedarian Project consisted of 111 children. With such a small sample, only large effects would have produced statistically significant findings. Thus, the absence of effects in some areas does not mean that the program did not affect these outcomes, but that the effects may have been too small to be detected with the sample size. A small sample also means that differences in baseline characteristics would also have to be very large to be statistically significant, making it more difficult to assess the comparability of the program and control groups.

Attrition. There was minimal attrition in the project. Although various reports produced about the Abecedarian Project detail the extent of attrition and the reasons for it, they do not describe the characteristics of those dropped from the Abecedarian evaluation. Although the sample size varied somewhat depending on the year of follow-up, about 90 to 104 of the randomly assigned children have been included in most analyses. At age twenty-one, 104 children were included in the analysis, representing an attrition rate of just 6 percent (assuming a base of 111 children).[42] This was remarkably little attrition, especially compared to other projects, such as the evaluation of the Early Head Start program, which had an attrition rate exceeding 30 percent for most outcomes after just two years.

The UNC team noted that the subjects lost to attrition did not differ from the original sample in terms of any entry level characteristics, but they did not specify the characteristics examined. Moreover, they based this comparison not on the original sample of 122, but on the already reduced sample of 111. Given the low rate of attrition (even using the 122 child sample as a base), attrition-related bias is not likely to be a serious concern, with one possible exception. Seven of the eight mothers that refused to participate after being randomly assigned were in the program group (and thus represented over 10 percent of that group). To the extent that their interest in being with their children also affected their children's outcomes, this disproportionate loss may have biased the findings.

Data collection. The data collection relied on a wide range of tests and survey results. The data sources are appropriate for the questions being studied and are relatively complete. The use of administrative data, however, was limited. The confidence surrounding some of the survey findings, particularly those dealing with employment, welfare use, and crime could have been strengthened by obtaining data from various administrative services, such as Unemployment Insurance records for employment and earnings.

Measurement issues. Most outcomes were measured using widely recognized measures, such as various IQ and achievement tests (such as the Wechsler Preschool and Primary Scale of

[42]Even adding the eleven children lost due to the randomization problem described above results in an attrition rate of just 15 percent (18 children of 122 randomly assigned).

Intelligence), school records, and periodic surveys.

Generalizability. The Abecedarian Project was composed almost exclusively of black children, despite the fact that black children were a distinct minority in the area. For example, 80 percent of the children in the school district were white and only 14 percent were black. According to the UNC team:

> It is important to note that the Abecedarian Project was located in a generally affluent college town where the vast majority of families were well-educated. Thus, there were relatively few families in the local population who would be considered socioeconomically at-risk. Those who did fit this category tended to be of African-American descent.[43]

There were also a wide range of public and private services for those in need, including those in the control group, and the schools to which many of the children went were good.

The elementary and secondary schools that the Abecedarian children subsequently attended tended to be better than those of many other early childhood interventions. This is an important point, because some researchers have hypothesized that the subsequent schooling experiences of children may affect the permanency of gains. For example, University of California, Los Angeles economists Janet Currie and Duncan Thomas estimated the persistence of Head Start gains by examining how the effects varied by the age of the child (see chapter 5).[44] They found that both white and black children who participated in Head Start had statistically significant gains of nearly seven percentile points on the Peabody Picture Vocabulary Test. By age ten, the gains for whites were about five percentage points, but the gains for blacks had faded out and were no longer evident. They hypothesized that the apparent impact of Head Start may have depended on the child's home background and the quality of schools attended after Head Start graduation.

Thus, the findings may generalize primarily to localities where poor children are in educationally well-off communities. Of even greater importance is that the Abecedarian children were chosen because—beyond just being poor—they were at risk of intellectual or academic failure.

In addition, the study was conducted over thirty years ago and may not be applicable to

[43]Craig T. Ramey, Frances A. Campbell, Margaret Burchinal, Martie L. Skinner, David M. Gardner, and Sharon L. Ramey, "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers," *Applied Developmental Science* 4, no. 1 (January 2000): 4, http://web.pdx.edu/~stipakb/download/PA555/EarlyChildhoodEducStudy.pdf (accessed June 28, 2010).

[44]Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" *American Economic Review* 85, no. 3 (June 1995): 341–364.

the current early education environment.

Replication. Despite major efforts, the project has not been replicated successfully. As Jonathan Crane, director of the National Center for Research on Social Programs, cautions:

> The program has not been formally upscaled or replicated. Because of the lack of replication, there is no information on the relationship between effect size and implementation fidelity or site experience. The most important reason for pause is that similar early intervention programs have not had consistent long-term effects on cognitive test scores . . . It is possible that the Abecedarian Project is simply one of two random outliers.[45]

Indeed, the Infant Health and Development Project (IHDP) (see chapter 16), which was modeled after the Abecedarian Project and conducted by an independent research firm, failed to achieve long-term gains in IQ or test scores.[46] There were, however, important differences in the target population and duration of services between the two programs, which could account for some of the difference in impacts. For example, the Abecedarian Project targeted low-income/high-risk families, whereas the IHDP was aimed at low-birthweight children. In addition, the Abecedarian Project provided services from birth to five years (and beyond, for some children), whereas IHDP services were limited to children from birth to age three. (Other replications could include the Milwaukee Project (although conducted earlier; see chapter 17) and CARE; however, because of their methodological flaws, we do not consider these valid replications.)

Evaluator's description of findings. Staff of the Frank Porter Graham Child Development Institute (FPG) has reported their findings with full confidence in the success of their program. A press release from the FPG Child Development Institute declares: "The importance of high quality, educational childcare from early infancy is now clear."[47] Although the findings may be promising, they are certainly not definitive, and do not provide a "clear" indication that high quality care is effective. In fact, this is but one study in one Southeastern town including mainly black children and thus cannot be generalized widely, especially in light of the disappointing findings of the IHDP and other early childhood interventions.

Age of entry is another important, but unresolved issue. As the authors themselves

---

[45]Jonathan Crane, "Building on Success," in *Social Programs That Work*, ed. Jonathan Crane (New York: Russell Sage Foundation, 1998), 22. The other outlier is the Milwaukee Project (see chapter 17).

[46]There were some positive effects for the subgroup of heavier low-birthweight infants, but these too showed a pattern of fading as the children grew older.

[47]Frank Porter Graham Child Development Institute, University of North Carolina, Chapel Hill, *Early Learning, Later Success: The Abecedarian Study* (Chapel Hill, NC: Frank Porter Child Development Institute, 1999), 2, http://www.fpg.unc.edu/~abc/ells-04.pdf (accessed June 29, 2010).

concede, further investigation is necessary: "Given that age of entry was confounded with other factors such as duration of treatment in this study, firm conclusions concerning the importance of beginning intervention in infancy cannot be drawn, but the results strongly support the further investigation of this factor in early intervention studies."[48] Thus, we cannot be sure that the effects were achieved because the project started in early infancy or because it lasted as long as it did, or for some other reason. It could be that participation in the later preschool years was more important (a finding that seems possibly consistent with some of the High/Scope Perry Preschool findings).

Evaluator's independence. On the one hand, some of the evaluators, including Ramey and Campbell, were closely associated with the FPG Child Development Institute. On the other hand, their findings have been published many times in peer-reviewed journals. Additionally, Masse and Barnett conducted independent analyses of the data in the course of their benefit-cost analysis.

Statistical significance/confidence intervals. Statistical significance was measured and reported at the 5 percent level.

Effect sizes. Effect sizes were calculated and reported, using two different calculation methods. Under the first method, the difference in the mean scores between the treatment group and the control group was divided by the pooled sample standard deviation (SD=6.30); and under the second method, the difference was divided by the standard deviation of the Wechsler IQ tests (SD=15), resulting in a more conservative estimate. At age four (the midpoint of the treatment period), the reported effect size for cognitive test scores was 1.75 SD using the first method and 0.74 SD using the second method. At age fifteen (the midpoint of the follow-up period), the reported effect size for cognitive scores was 0.87 using the first method and 0.37 using the second method. These effect sizes were based on statistically significant impacts.

The UNC team describe these effects as meaningful, noting, "Educators consider effect sizes of 0.25 or greater to have practical significance (Cohen 1977). . . . even the more conservative estimates are in the range considered educationally meaningful, and both estimates for the period of time during the treatment are considered large."[49] Based on traditional demarcations, this conclusion seems reasonable. (See Appendix 1 for a further discussion of effect sizes and their interpretation.)

---

[48]Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 239.

[49]Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 236.

Sustained effects. The Abecedarian evaluation examined impacts through age twenty-one.

Benefit-cost analysis. A benefit-cost analysis was conducted, but not by the principal investigators of the study. The analysis, however, was from the perspective of society as a whole. It did not include the participant or taxpayer perspectives.

Cost-effectiveness analysis. Apparently not performed.

Commentary

Craig T. Ramey and Frances A. Campbell*

We appreciate the invitation to comment on the chapter summarizing and interpreting the outcomes from the Abecedarian Project. Brief comments on key criticisms follow.

Concerns about bias due to differential refusal of group assignment

The writers raise questions about initial random assignment and the fact that "over 10%" of the treated sample refused their random assignments.

The initial randomization procedures for the Abecedarian study met accepted scientific standards. The sampling plan involved sampling with replacement. Once a sufficiently large pool of potential participants was identified, families within it were randomly assigned to either the treatment or the control group. While in the ideal world, everyone who met recruitment criteria would accept their randomly assigned group, this rarely happens. The accepted practice in a case of refusal is to sample for cases to replace slots in the group where the refusal occurred, not for double that number when there would be no comparable slot in the other group. This explains why, if more persons offered treatment chose not to accept, more families would necessarily have been assigned to the treated group.

Selection bias would result if parents who refused treatment differed from those who remained in the program in ways that interacted with treatment to affect children's outcomes. If this were the case, significant differences between parents who refused to participate and those who accepted their assignment would be found, and treatment would significantly interact with the relevant parental characteristic, thus causing a differential effect of treatment on child outcomes. We have not found systematic differences between refusing parents and other

*Craig T. Ramey and Frances A. Campbell are both affiliated with the Frank Porter Graham Child Development Institute at the University of North Carolina. Craig T. Ramey was the principal investigator of the original study, and Frances A. Campbell was the principal investigator of the age twenty-one and later follow-up studies.

parents.[1] Neither have we found differences by family characteristic interactions.[2] There is no indication that initial bias artificially inflated the treatment effect.

The assertion that any treatment gains in the program "appear to have been concentrated among the subgroup of children whose mothers had IQs below 70 at the time of entry to the study"

This statement could only be true had the Abecedarian data shown a treatment by maternal IQ interaction when predicting child IQ. We have tested for such interactions and not found them.[3] In contrast, our results demonstrate that treatment and maternal IQ contribute independently to children's cognitive growth. The writers of this chapter may have been confused by the fact that the size of the treatment/control IQ difference among the few children of such mothers has been described on at least two occasions.[4] In both instances, the information provided was descriptive. Neither article reports an analysis in which the model tested included a term for the interaction of maternal IQ and treatment. In analyses where this interaction has been tested, it was not similarly in the treated and control group children and the effects of treatment do not vary by maternal IQ. Therefore, the statement that treatment benefits "appear to have been concentrated among the subgroup of children whose mothers had IQs below 70 at the time of entry to the study" reflects a basic misunderstanding of the body of results published over the years in the refereed literature.

Concerns that the effect of treatment on cognitive function was small

---

[1]Sandra L. Martin, Craig T. Ramey, and Sharon Ramey, "The Prevention of Intellectual Impairment in Children of Impoverished Families: Findings of a Randomized Trial of Educational Daycare," *American Journal of Public Health* 80, no. 7 (July 1990): 844–47.

[2]Craig T. Ramey, Keith O. Yeates, and Elizabeth J. Short, "The Plasticity of Intellectual Development: Insights from Preventive Intervention," *Child Development* 55, no. 5 (October 1984): 1913–1925; and Margaret R. Burchinal, Frances A. Campbell, Donna M. Bryant, Barbara M. Wasik, and Craig T. Ramey, "Early Intervention and Mediating Processes in Intellectual Development Among Low-Income African American Children," *Child Development* 68, no. 5 (October 1997): 935–954.

[3]Craig T. Ramey, Keith O. Yeates, and Elizabeth J. Short, "The Plasticity of Intellectual Development: Insights from Preventive Intervention," *Child Development* 55, no. 5 (October 1984): 1913–1925; and Margaret R. Burchinal, Frances A. Campbell, Donna M. Bryant, Barbara M. Wasik, and Craig T. Ramey, "Early Intervention and Mediating Processes in Intellectual Development Among Low-Income African American Children," *Child Development* 68, no. 5 (October 1997): 935–954.

[4]Sandra L. Martin, Craig T. Ramey, and Sharon Ramey, "The Prevention of Intellectual Impairment in Children of Impoverished Families: Findings of a Randomized Trial of Educational Daycare," *American Journal of Public Health* 80, no. 7 (July 1990): 844–47; and Frances A. Campbell and Craig T. Ramey, "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention," *American Educational Research Journal* 32, no. 4 (Winter 1995): 743–772.

The most widely accepted guidelines for evaluating effect sizes[5] suggest that effect sizes (d: computed as a difference in group means divided by the relevant standard deviation) should be regarded as small when d=.20, moderate when d=.50, and large when d is .80 or higher. Based on Cohen's definitions, the treatment/control group effect sizes for cognitive outcomes qualify as large when the pooled sample standard deviation is used to compute it (the most appropriate method since the findings generalize specifically to a high-risk sample) and small to moderate when the test standard deviation is used (one that is based on a sample that includes high and low risk individuals).[6] Furthermore, it is noteworthy that, through age twenty-one, the intellectual test score functions for the treated and control groups do not converge, but rather change in parallel after the early childhood years. Longitudinal analyses (repeated-measures design) conducted at several points over the course of the study always show a significant difference between the intellectual test scores of the treated and control groups.[7] In addition, the preschool treatment/control group differences in intellectual test scores has been examined separately at thirteen time points between the ages of two and twenty-one years. Statistically significant group differences were seen at eleven of the thirteen isolated points.

Far more important is the fact that in young adulthood the treated group demonstrated important real-life benefits in terms of better scores on objective reading and mathematics tests, reductions in early childbearing, more years of education attained, and employment in higher level jobs (discussed in more detail below).

Concerns that treatment effects were not linked to educational or employment outcomes in early adulthood

This assertion appears to be based on the finding of no treatment/control differences in high school graduation or overall employment rates in young adulthood. However, this is

[5]Jacob Cohen, *Statistical Power Analysis,* 2nd ed. (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 1988).

[6]Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret R. Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 231–244.

[7]Craig T. Ramey and Frances A. Campbell, "Preventive Education for High-Risk Children: Cognitive Consequences of the Carolina Abecedarian Project," *American Journal of Mental Deficiency* 88, no. 5 (March 1984): 515–523; Craig T. Ramey, and Frances A. Campbell, "Poverty, Early Childhood Education and Academic Competence: The Abecedarian experiment," in *Children in Poverty: Child Development and Public Policy,* ed. Aletha C. Houston (New York: Cambridge University Press, 1991), 190-221; Frances A. Campbell and Craig T. Ramey, "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families, *Child Development* 65, no. 2 (April 1994): 687–689; Frances A. Campbell and Craig T. Ramey, "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention," *American Educational Research Journal* 32, no. 4 (Winter 1995): 743–772; and Frances A. Campbell, Elizabeth P. Pungello, Shari Miller-Johnson, Margaret R. Burchinal, and Craig T. Ramey, "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment," *Developmental Psychology* 37, no. 2 (March 2001): 231–244.

misleading because it ignores several significant outcomes found in the age twenty-one follow-up study.

For example, we learned at the age twenty-one assessment that, compared with the control group, the individuals in the treatment group had significantly more years of education (a benefit with a significant treatment by gender interaction, indicating it was largely limited to females) and were significantly more likely to hold a job ranked four or higher on the Hollingshead scale (a benefit not limited to females). We assume that, in pointing out the gender difference in years of education attained, the writers did not intend to imply that this in any way diminished the importance of this post-secondary education gain. More four-year college or university (the gateway to professional jobs) is a benefit that did not show a significant treatment by gender interaction, and thus, one that could be applied both to males and females.

As published in one of the peer-reviewed articles reporting the young adult findings[8] the individuals in the treatment group were significantly more likely than the individuals in the control group to be engaged in either higher education or skilled/professional level employment. Thus, the fact that, at age twenty-one, for the preschool treatment and control groups, there were similar rates of high school graduation and of employment at any level, does not equate to there being no group differences in post-secondary educational attainment or employment in higher-level jobs.

Concerns that cost-benefit figures are inflated

A noted economist, Dr. Steve Barnett, of the National Institute for Early Education Research, and his student, Dr. Leonard Masse, have conducted an independent cost-benefit analysis of the Abecedarian Project. However, we would like to note that the Nobel Prize-winning economist, Dr. James Heckman of the University of Chicago, has examined Masse and Barnett's work and did not raise these concerns.[9] Instead, Dr. Heckman argued that cost-benefit analyses to date, including this one of the Abecedarian study along with others, indicate that early childhood programs offer one of the best opportunities for our society to further economic growth and for children from low-income families eventually to achieve betters lives.

Conclusions

In summary, the authors have raised a number of concerns about the study's methods and also question whether the findings support the value of early childhood education. In general,

[8]Frances A. Campbell, Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson, "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project," *Applied Developmental Science* 6, no. 1 (January 2002): 42–57.

[9]James Heckman, "What's Good Early Childhood Education Worth and Why?" (discussant, Biennial Meeting of the Society for Research in Child Development, Tampa, FL, April 26, 2003).

they have highlighted non-significant treatment/control differences found (e.g., similar rates of high school graduation, overall employment, and lawbreaking at age twenty-one) while downplaying a number of significant findings (e.g., better reading and math achievement scores, higher rates of attendance at four-year college or universities, employment in higher-level jobs, and reductions in early childbearing). That a well-controlled experimental study of an early childhood program could demonstrate such important real-life benefits years after treatment ended offers strong support for providing early learning supports for children who grow up in poverty.

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 or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter2.html.