SCHOOL of PUBLIC POLICY

7
Early Head Start Research and Evaluation 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/chapter7.html.

7
Early Head Start Research and Evaluation Project

The Early Head Start program began in 1995 as a comprehensive, two-generation program intended to enhance children's development and help parents educate their young children. As of 2009, there were more than 650 Early Head Start programs serving more than 66,000 low-income children and their families at a cost of about 10,700 per child.[1] Early Head Start serves pregnant women and low-income families with infants and toddlers. The program was developed partially in response to a 1994 report from the Carnegie Corporation that described a "quiet crisis" facing families with infants and toddlers, with many low-income children starting life in subpar environments and without sufficient interaction with caring adults.[2] Shortly thereafter, the Advisory Committee on Head Start Quality and Expansion recommended that Head Start programs "address the fragmentation of services by forging new partnerships, and expand services in a number of ways, including serving more families with infants and toddlers."[3]

John M. Love and Ellen Eliason Kisker of Mathematic Policy Research and Jeanne Brooks-Gunn of Columbia University led the national evaluation team (the "MPR team"). The first phase of the Early Head Start Research and Evaluation Project used data from 1996–1998 and followed children from birth through their first three years.[4] The research design was based

---

[1]U.S. Department of Health and Human Services, Administration for Children and Families, Office of Head Start, *Head Start Fact Sheet: Fiscal Year 2010* (Washington, DC: U.S. Department of Health and Human Services, 2010), http://www.acf.hhs.gov/programs/ohs/about/fy2010.html (accessed July 1, 2010).

[2]Carnegie Corporation of New York, *Starting Points: Meeting the Needs of Our Youngest Children* (New York: Carnegie Corporation of New York, 1994).

[3]Ellen Eliason Kisker, John M. Love, and Helen Raikes, *Leading the Way: Characteristics and Early Experiences of Selected Early Head Start Programs. Volume 1: Cross-Site Perspectives* (Washington, DC: U.S. Department of Health and Human Services, December 1999), 4.

[4]The second phase of the project followed the children from the time they left the program until they entered kindergarten. Data collection was slated to be completed in 2005, but, as of publication, the final evaluation has yet to be released. ACF funded a third phase of the program to follow the children through grade five. Data collection began in 2007 and was scheduled to be completed in 2010. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, "Early Head Start Research

on random assignment, which appears to have been carried out well and with a large sample size. The evaluation reported a pattern of statistically significant, but small, impacts across a range of child and parent outcomes with some important exceptions. The effect sizes for virtually all outcomes, however, fell below levels that have traditionally been considered educationally meaningful. Moreover, a high attrition rate is likely to undermine any longer-term follow-up, leaving unclear the program's effects with respect to many important outcomes, such as school readiness.

Program Design

Program group. At the time of this evaluation, Early Head Start was aimed at pregnant women and families with children under three years of age who met Head Start's income eligibility guidelines.[5] In the research sites, for evaluation purposes, the age limit for entry into the program was lowered to one year (or younger). Most families enrolled before the child reached six months and about one-fourth enrolled before the child was born.

The characteristics of Early Head Start families in the research sites varied considerably at the time of enrollment: On average, 40 percent were two-parent families (with a range among sites from 9 to 74 percent); one-third were headed by a teen parent (with a range from 12 to 84 percent); over one-third were white (one-third were black and one-fourth were Hispanic); just over half of the primary caregivers had a high school degree (with a range from 12 to 76 percent); and less than one-fourth were employed (with a range from 11 to 44 percent).[6]

Services. At the time of the evaluation, the Early Head Start program offered a wide range of services and gave grantees considerable flexibility in designing services, including child development, parenting education, child care, home visiting, and family support services. The primary objective of the program was to strengthen parenting skills and parent-child relationships. Because many past adult-focused programs (such as the CCDP, see chapter 3) had failed, all three Early Head Start approaches also placed a heavy emphasis on child development services.

---

and Evaluation Project: Evaluation Team," http://www.acf.hhs.gov/programs/opre/ehs/ehs_resrch/ehs_overview.html#team (accessed July 1, 2010).

[5]Ninety percent of families must have incomes at or below the poverty level or be eligible for public assistance and programs must make at least 10 percent of spaces available for children with disabilities.

[6]Kisker, Love, and Raikes, 1999, 25.

Three program approaches were tested: center-based, home-based, and mixed.[7] Center-based programs provided most services in a center and were supplemented with some home visits. These programs generally offered full-time child care with low child-to-staff ratios (4:1 or lower) and group sizes of no more than eight children. Home-based programs emphasized weekly home visits and included some group socialization. Some programs offered a mixed approach depending on the needs of the family.

The Evaluation. The national evaluation of Early Head Start was carried out by Mathematica Policy Research, Inc. and the National Center for Children and Families at Columbia University's Teachers College in collaboration with fifteen local research groups.[8] It is a random assignment evaluation in seventeen sites. The sites were selected to be representative of the "programs funded in 1995 and 1996, including their program approaches and family demographic characteristics."[9]

Between July 1996 and September 1998, about 3,000 families were randomly assigned to either a program or a control group. The data are based on periodic parent interviews, direct child assessments, and videotaped parent-child interactions.[10] As originally planned, the last follow-up of children was to be conducted when the children were three years old. Subsequently, the U.S. Department of Health and Human Services funded additional assessments of the children and interviews with their parents in spring or summer just prior to their kindergarten enrollment.[11]

---

[7]The Head Start Program Performance Standards define program options that programs provide to families. Programs may offer one or more options to families, including (1) a home-based option, (2) center-based option, (3) a combination option in which families receive a prescribed number of home visits and center-based experiences, and (4) locally designed options. Because a single program may offer families multiple options, for purposes of the research, the researchers characterized program approach according to the options offered families, (1) home-based (all families in a home-based option), (2) center-based (all families in a center-based option), or (3) mixed (some families receive the home-based and some the center-based option or some or all receive the combination option).

[8]The initial cost of the evaluation was $21 million over seven years, but this was based on a thirty-six month follow-up. The extension of the follow-up, described below, will increase costs further.

[9]U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Commissioner's Office of Research and Evaluation and Head Start Bureau, *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families, Summary Report* (Washington, DC: U.S. Department of Health and Human Services, January 2001), 2, http://www.mathematica-mpr.com/PDFs/buildsumm.pdf (accessed December 30, 2002).

[10]Parent interviews that focused on services used were conducted six, fifteen, and twenty-six months after enrollment. Interviews that included more age-related information were conducted along with the Bayley Scales of Infant Development Mental Development Index and videotaped interactions at fourteen, twenty-four, and thirty-six months of age.

[11]John Love, Mathematica Policy Research, e-mail message to Peter Germanis, February 16, 2001.

Major Findings

The most recent findings of Early Head Start are for children at age three.[12] At age three, the Early Head Start children were functioning somewhat better than their control group peers on some cognitive, language, and social-emotional development measures. The MPR team concludes:

> The Early Head Start research programs stimulated better outcomes along a range of dimensions (with children, parents, and home environments) by the time children's eligibility ended at age 3. Overall impacts were modest, with effect sizes in the 10 to 20 percent range, although impacts were considerably larger for some subgroups, with some effect sizes in the 20 to 50 percent range. The overall pattern of favorable impacts is promising, particularly since some of the outcomes that the programs improved are important predictors of later school achievement and family functioning.[13]

The effect sizes for virtually all outcomes, however, fell below levels that have traditionally been considered educationally meaningful.

Cognitive. At age two, Early Head Start children achieved a statistically significant 2 point gain on the Bayley Mental Development Index (MDI), a standardized assessment of infant and toddler cognitive development (90.1 vs. 88.1 for control group children). Perhaps more significant, a smaller percentage of Early Head Start children (33.6 percent compared to 40.2 percent for the control group) fell in the "at-risk" range of developmental functioning (below 85 on the Bayley MDI), an effect size of -0.14 SD.

Children in the program group also had significantly higher scores on the MacArthur Communicative Development Inventory (CDI) for vocabulary (56.3 vs. 53.9, an effect size of 0.11 SD) and sentence complexity (8.6 vs 7.7, an effect size of 0.11 SD). There was no statistically significant difference on the CDI for combining words.

At age three, Early Head Start children maintained this statistically significant 2 point gain on the Bayley MDI (91.4 vs. 89.9 for control group children). In addition, a smaller percentage of Early Head Start children (27.3 percent compared to 32.0 percent for the control

[12]Unless otherwise indicated, all findings from: John Love, Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Diane Paulsell, Kimberly Boller, Jill Constantine, Cheri Vogel, Allison Sidle Fuligni, and Christy Brady-Smith, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Volume 1: Final Technical Report* (Washington, DC: U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Commissioner's Office of Research and Evaluation and Head Start Bureau, June 2002), http://www.mathematica-mpr.com/PDFs/ehsfinalvol1.pdf (accessed December 30, 2002).

[13]Love et al., 2002, xxv.

group) fell in the "at-risk" range of developmental functioning (statistically significant at the 10 percent level). But these impacts were smaller than at age two, with the effect sizes just 0.12 SD and -0.10 SD. Program children had significantly higher scores on the Peabody Picture Vocabulary Test (PPVT-III) than children in the control group but, again, the effect size was small (0.13 SD).

There are several other reasons why these findings should be considered disappointing. First, the effect sizes for all outcomes are small. Most researchers consider only those effect sizes of 0.2 or larger to be educationally meaningful. (See Appendix 1 for a further discussion of effect sizes and their interpretation.) Only one of the sixty-four main impacts reported in the Early Head Start evaluation meet this criterion, and most are quite a bit smaller.[14]

Many early childhood programs have reported achieving long-lasting improvements in school performance (such as grade retention) without achieving lasting IQ or other achievement gains. However, the relatively short follow-up period of the evaluation will preclude an assessment of such outcomes. Thus, the practical significance of improvements in children's social-emotional development or the parent's child-rearing skills, to the extent they persist, will remain uncertain.

School readiness/performance. Apparently data either were not collected or reported.

Socioemotional development. At age two, program children were significantly less likely to be rated as having aggressive behavior on the Child Behavior Checklist, but the effect size was small (-0.10 SD), and the measure relied on parental reports. There were no statistically significant differences on five other measures of socioemotional development.

At age three, Early Head Start children continued to be rated as less aggressive and were also found to engage their parents more. During semistructured play, they showed less negativity toward their parents and were more attentive. There were no statistically significant differences on five other socioemotional development measures.

Health. There were no statistically significant effects on direct health measures, but by age three, Early Head Start children were slightly more likely to be immunized (99 percent vs. 98 percent). In addition, they were less likely to have been hospitalized for an accident and or injury in the third year of follow-up (0.4 percent vs. 1.6 percent).

Behavior. See socioemotional development.

[14]While some subgroup effect sizes were a bit larger, they too were small. For example, the effect size for the Bayley MDI in center-based programs was 0.22 SD and 0.28 SD in early-implemented mixed-approach programs (where the impact on reducing the percentage scoring below 85 on the PPVT-III had an effect size of -0.34 SD).

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. Early Head Start generated a number of apparently positive impacts on measures of parenting and the home environment.

Parenting knowledge and discipline strategies. When the children were age two, parents of children in the program group had significantly higher knowledge of infant-toddler development, although the difference in the average score was only 0.1, and also were more likely to suggest positive discipline strategies. There were no statistically significant differences on four other measures of parental knowledge and discipline strategies. When the children were age three, parents of children in the program group were less likely to have spanked their child in the past week (46.7 percent vs. 53.8 percent, an effect size of 0.14 SD) and were less likely to suggest negative discipline strategies. There were no statistically significant differences on four other measures of parental knowledge and discipline strategies.

Parent's physical and mental health and family functioning. When the children were two, parents of children in the program group were less likely to have parental stress and less likely to have family conflict but the actual differences are extremely small. There were no statistically significant differences on four other measures of the parents' physical or mental health. When the children were three, there were no statistically significant differences on all six measures of the parents' physical and mental health.

Parent self-sufficiency. When their children were two, parents of children in the program group were more likely to have participated in an education or job training program (48.4 percent vs.43.7 percent, an effect size of 0.11 SD) and to have spent more time per week in an education or job training program (5.3 hours vs. 4.1 hours, an effect size of 0.15 SD). There were no statistically significant differences on five other measures of parent self-sufficiency. These findings persisted when the children were three.

Parenting behavior. When the children were two, parents of children in the program group were more likely to be supportive during parent-child play, more likely to read to their children every day, more likely to read to their children at night, more likely to set a regular bedtime, and scored higher on the Home Observation for Measurement of the Environment (HOME) index. There were no statistically significant differences on four other measures of parenting behavior.

When the children were three, parents of children in the program group were more likely to be supportive during parent-child play, more likely to read to their children every day, and

more likely to score higher on the HOME index. There were no statistically significant differences on thirteen other measures of parenting behavior.

In addition, the home environments of Early Head Start participants were more supportive and stimulating of development than were those of the control group. The largest impacts were observed for home-based and mixed-approach programs. However, the programs had few impacts on measures related to parental self-sufficiency. During the first twenty-six months, there were no statistically significant effects on welfare receipt or family income. When using a 10 percent level of significance, there was a modest gain in the percent ever employed over twenty-six months (86.8 percent vs. 83.4 percent).

Subgroup effects. The Early Head Start evaluation, in addition to examining the overall effects of the program, looked at its impact on specific demographic groups. In order to do so, the MPR team averaged site impacts across sites where there were at least ten program and ten control families in the subgroup. Five demographic factors were used to evaluate a family's risk level, including (1) being a single parent, (2) receiving public assistance, (3) being neither employed nor in school or job training, (4) being a teenage parent, (5) lacking a high school diploma or GED. Families with zero, one, or two risk factors were considered low-risk; those with three risk factors were considered medium-risk; and those with four or five risk factors were considered high-risk.

The MPR team concludes that Early Head Start benefitted all types of children and families, but they note that the positive effects were most likely to reach families of moderate disadvantage. For families of a high or low level of disadvantage, the magnitude of the effect was much smaller, and in some cases negative. For example, on the Bayley MDI, low-risk children had an effect size of about 0.01, moderate-risk children had an effect size of about 0.25 and high-risk children had an effect size of about -0.14. Only the effect for moderate-risk children was statistically significant. There was no statistically significant impact on the percentage of children that scored less than 85 on the Bayley MDI. For the most part, subgroup effects, with the exception of parenting behavior, tended to be small and were rarely statistically significant.

For some subgroups, the impacts on child social-emotional development (including parenting behavior) were statistically significant and had modest (rather than small) effect sizes. Some of these measures, however, appeared to be quite similar and thus, seem to exaggerate the overall importance of the effects. For example, it does not seem surprising that both "negativity toward parent during parent-child semistructured play" and "supportiveness during parent-child semistructured play" were statistically significant in the overall findings. And, while some consider such effects to be encouraging, it is unclear how these small effects could have any real, long-term impact on children.

Nevertheless, one strength of the evaluation is the large number of outcomes examined for each subgroup and the use of parent surveys, tests, and independent observations to measure

them. But this is also a possible pitfall of the study, as it is almost always possible to find some positive impacts if one divides the treatment group into enough separate groups.

Benefit-cost findings. In 2009, the average cost of Early Head Start was approximately $10,700 per family per year (in 2009 dollars).[15] For the program to produce savings that exceed these costs, it would have to produce longer-term reductions in grade retention, special education placement, criminal activity, or other outcomes that produce offsetting savings. The Early Head Start evaluation will not follow the children for a sufficiently long enough period to determine whether such savings occur. However, the effect sizes reported so far are unlikely to result in benefits anywhere approaching the program's cost.

Overall Assessment

The Early Head Start evaluation was based on random assignment, which appears to have been carried out well and with a large sample size. However, a high attrition rate for some outcomes after three years is of some concern, particularly if the problem worsens as follow-up data collection is completed.

Program theory. Early Head Start programs were designed to "enhance children's development and health, strengthen family and community partnerships, and support the staff delivering new services to low-income families with pregnant women, infants, or toddlers."[16] Early Head Start programs could choose one of three approaches to implementing this two-generation program: the center-based option, the home-based option, or a mixed approach. There is no other specific theory detailed beside the general expectation that early childhood intervention programs promote school readiness and improve developmental outcomes for children.

Program implementation. Although there were some initial implementation difficulties, the MPR team concluded that the Early Head Start programs provided the intended services to nearly all families.[17] For example, 95 percent of the families in the program group received at least one service within sixteen months of enrollment. Most received at least one home visit (86.7 percent) and met with a case manager (85.3 percent), but smaller percentages participated in group parenting activities (67.3 percent) or placed their children in center-based care (42.6 percent).

---

[15]John Love, Mathematica Policy Research, e-mail message to Peter Germanis, November 24, 2001. This figure is based on the cost of the 664 Early Head Start programs funded in the Fall of 2001.

[16]U.S. Department of Health and Human Services, 2001, 1.

[17]Implementation studies were conducted during site visits in 1997 and 1999. These visits documented to the programmatic approaches, the services provided, and measured other aspects of program performance.

Note that these figures are for after sixteen months. Indeed, it appears that these services were not a regular part of the Early Head Start package at many sites. For example, while 86.7 percent of the families had a home visit, just 44.4 percent had a visit at least once a week during the six months after enrollment. Similarly, the percentage that met with a case manager at least once was 85.3 percent, but only 47.4 percent met with a case manager on a weekly basis during the first six months after enrollment. Control group families received many of the same services, but on a smaller scale. For example, while 95 percent of families in the program group received a key service within sixteen months, so did 75 percent of control group families.

Six of the programs reached "full implementation" by 1997 (termed "early implementers"), six more did so by Fall 1999 (termed "late implementers"), and five were still incompletely implemented by 1999. The failure of about two-thirds of the programs to reach "full implementation" within the first year may partially explain the relatively modest impacts reported.

The MPR team explored whether program effectiveness was a function of the level of implementation. They found that early, full implementation was associated with higher levels of program engagement among program group families.[18] Although this is not an experimental finding, it is suggestive.

Assessing the randomization. Between July 1996 and September 1998, a total of 3,001 families were randomly assigned to program (1,513) and control (1,488) groups, with about 150 to 200 families per site. Although the denial of services to nonparticipants is a common objection to random assignment, this was not an issue with the Early Head Start evaluation, as the program was new and only served a small portion of those who met the program's eligibility criteria.

No serious problems in implementing random assignment were reported.[19] A comparison of baseline characteristics showed just three statistically significant differences (at the 0.10 level) of forty-seven tested, which is less than the five that would be expected by chance.[20] The MPR

[18]Love, et al., 2002, 138.

[19]In some sites, however, the MPR team altered the normal enrollment procedures: Although enrollment was proceeding, for some programs the research eligibility criteria made it harder to recruit families, and the need to recruit twice as many families (to allow for control group assignment) made it harder to meet the deadline for full enrollment. If the applicants brought into the program were not equivalent to those who were effectively displaced when assigned to the control group, the impacts on those who were in the program may not correspond to the impacts on those who would have been in the program in the absence of the evaluation. *See* Kisker, Love, and Raikes, 1999, 16.

[20]For a summary of impacts expressed on a per applicant basis, *see* John M. Love, Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Kimberly Boller, Diane Paulsell, Alison Sidle Fuligni, and Lisa S. Berlin, *Building Their Futures: How Early Head Start Programs are Enhancing the Lives of Infants and Toddlers in*

team concludes that the program and control groups were equivalent.

There were two other issues related to random assignment. First, a small number of those assigned to the program group did not participate in Early Head Start. To estimate impacts for those who actually participated, the MPR team divided the impact per eligible applicant by the group participation rate, which increased the estimated impacts of the program.[21] Most other evaluators report their findings including all those who were assigned to the program group, regardless of whether they participated.[22] Since participation rates were 91 percent, this procedure had a relatively small effect on estimated impacts, increasing estimated impacts by about 10 percent.[23] Nevertheless, this is not the conventional approach in presenting the findings in other evaluations of early childhood education programs.

In addition, there were a small number of "crossovers," or control group families that participated in Early Head Start. Such "crossovers," however, accounted for only 0.7 percent of the control group and would be expected to exert a small downward bias on the estimated impacts, assuming the "crossover" cases benefitted from the program. The MPR team statistically controlled for this problem.

Assessing statistical controls in experimental and nonexperimental evaluations. The evaluation was based on random assignment, so selection bias should not be a serious problem. Moreover, the MPR team used regression analysis to control for twenty-eight baseline child and family characteristics that remained after random assignment, presumably improving the precision of the estimated impacts.

Sample size. The sample of 3,000 families is large enough to detect most meaningful impacts overall and to provide estimates for specific sites[24] and subgroups.

Attrition. When the children were age three, 70 percent of the parents completed the parent interview and 55 percent of the children completed the Bayley MDI and videotaped

---

Low-Income Families. Volume II: Technical Report Appendices (Washington, DC: U.S. Department of Health and Human Services, January 2001), D.65–81.

[21]The definition of participation was liberal: more than one home visit, more than one meeting with a case manager, child enrolled in an Early Head Start center for at least two weeks, or participated in a group activity.

[22]For example, see the Head Start Impact Study, this volume, chapter 13.

[23]Love et al., 2001, vol 2., *Appendices*.

[24]Although the MPR team did not make site-specific estimates, the average sample size of about 175 families per site exceeds the total sample size of many model programs, such as the Perry Preschool Project (123) and the Abecedarian Project (122).

assessments.[25] Response rates in some sites were considerably below these averages. The researchers report that, "Nonrespondents to the surveys and assessments tend to be somewhat more disadvantaged than respondents on a few dimensions, according to analyses of baseline characteristics."[26] They also noted that the response rate was about 2 to 6 percentage points higher for the program group than the control group, but after examining the baseline characteristics of the two groups, they conclude that "this differential did not result in any attrition bias."[27] To minimize potential bias, the MPR team controlled for differences in the baseline characteristics of program and control group respondents. Of course, there may have been unmeasured differences between the two groups, so attrition-related bias cannot be ruled out. The high level of attrition after just two to three years is of particular concern, especially surrounding the Bayley MDI and videotaped assessments. Moreover, any further follow-up surveys or assessments may have response rates so low as to undermine the credibility of the longer-term findings.

Data collection. The data collection relied on a wide range of tests and survey results. The data sources were appropriate for the questions being studied and were relatively complete.

Measurement issues. The MPR team used many nationally recognized tests and assessment tools.

A possible source of problems is that many of the impacts related to parental and child well-being were based on parent surveys. But as a report by the Manpower Demonstration Research Corporation cautions: "Program group parents' perceptions of their children may be influenced by their experiences in the program, differences between their reports and those of control group members may reflect only those differences in perception than differences in children's actual functioning."[28] It is not clear, however, whether the results in the Early Head Start evaluation would have overstated or understated program impacts. In general, "other studies involving maternally reported behavior problems at this age indicate that such reports correspond to clinically detected problems and may be predictive of longer-term adverse outcomes."[29] Many

[25]Love et al., 2002, 48.

[26]U.S. Department of Health and Human Services, 2001, 3.

[27]Love et al., 2002, 48.

[28]Pamela A. Morris, Aletha C. Huston, Greg J. Duncan, Danielle A. Crosby, and Johannes M. Bos, *How Welfare and Work Policies Affect Children: A Synthesis of Research* (New York: Manpower Demonstration Research Corporation, March 2001), 16.

[29]The Infant Health Development Program, "Enhancing the Outcomes of Low-Birth-Weight, Premature Infants: A Multisite Randomized Trial," *Journal of the American Medical Association* 263, no. 22 (June 13, 1990): 3041.

other impacts, however, were based on observational measures.

Generalizability. The generalizability of the Early Head Start evaluation is very strong. The seventeen research sites were selected to represent the major regions of the country, to include urban and rural areas, and to test the different programmatic approaches permitted by the program. The programs in these sites captured the national variation in terms of center-size, racial distribution of children, family type, and other characteristics. In addition, the programs appear to have been implemented well and were selected from existing programs, without all the costs of special resources and staff associated with model programs.

It will not be possible to test all variations in each type of geographic or residential area. For example, all three programs in the South are center-based, so there will be no information on the effectiveness of the home-based model in this region. Nevertheless, the findings should be broadly representative of program effects nationally.

Other factors—mainly changes in program approach and eligibility criteria—affect the generalizability of the findings. At the time of initial funding, the programs were divided relatively evenly among the three programmatic approaches, but by 1999, most had adopted a mixed approach (see table 1).[30] The shift in focus arose in part due to changing family needs, but the changes make it more difficult to isolate the impact of a particular program approach.

Table 1. Early Head Start: Changes in Program Approaches Over Time in Seventeen Sites

	Home-based approach	Center-based approach	Mixed approach
When funded	5	5	7
Fall 1997	7	4	6
Fall 1999	2	4	11

Source: U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Commissioner's Office of Research and Evaluation and Head Start Bureau, *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families, Summary Report* (Washington, DC: U.S. Department of Health and Human Services, January 2001), 19, http://www.mathematica-mpr.com/PDFs/buildingvol1.pdf (accessed December 30, 2002).

The research sites imposed a requirement that children be under age one at initial application. If the magnitude of impacts are related to duration of participation, then the evaluation will exaggerate the impact of the basic Early Head Start program, which admits many older children. Many mothers prefer informal arrangements for infants and may not enroll children in center-based programs until age two or three.

[30]Love et al., 2002, 19.

Replication. There have been no random assignment replications of this program.

Evaluator's description of findings. In what seems an optimistic statement, the MPR team concludes that these small, early effects are promising:

> The initial impacts emerging from the evaluation of the new Early Head Start programs are promising. The pattern of modest but significant impacts across a wide range of child and parent outcomes at a point about two-thirds of the way through children's Early Head Start program experience suggests that the programs are reducing the risk that children will experience poor outcomes later on.[31]

This statement is reminiscent of similar conclusions reached by the Infant Health and Development Project (IHDP), when initial positive (but much larger) effects for children at age three led the MPR team to conclude that the "comprehensive and intensive early intervention program shows substantive promise of decreasing the number of LBW [low birthweight] premature infants at risk for later developmental disability."[32] Subsequent follow-ups through age eighteen showed that many of the initial effects disappeared.[33]

Some of the conclusions in the report seem to go beyond an objective reading of the evaluation's results. For example, evidence of reductions in parenting stress, family conflict, and other related behaviors led the MPR team to conclude that "Early Head Start programs may be helping to break a cycle of stress, conflict, poor coping strategies and punitive discipline sometimes reported in studies of low-income families."[34] Parent reports of spanking or stress based on one (possibly two) observations may not be typical of long-term behaviors. Judgments about breaking a "cycle" of behaviors surely require longer-term follow-up with confirmation from more objective, collateral sources.

Evaluator's independence. Early Head Start is being evaluated under contract with the U.S. government by an independent team, headed by Mathematica Policy Research and researchers at Columbia University, in collaboration with fifteen local research teams.

---

[31]U.S. Department of Health and Human Services, 2001, 12.

[32]The Infant Health and Development Program, "Enhancing the Outcomes of Low-Birth-Weight, Premature Infants: A Multisite Randomized Trial," *Journal of the American Medical Association* 263, no. 22 (June 13, 1990): 3041.

[33]Similarly, the CCDP found a similar pattern on many small positive effects at age two (including a 2-point difference on the Bayley MDI), all of which disappeared at age five. Early Head Start maintained a statistically significant finding at age three, but the CCDP did not.

[34]U.S. Department of Health and Human Services, 2001, 12.

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

Effect sizes. For statistically significant impacts, most reported effect sizes fell in the range of 0.10 to 0.20 SD. The MPR team identifies these effects as "favorable." This reflects a move away from the traditional demarcations that would not consider such effects relevant to policy. For some subgroups, reported effect sizes fell in the range of 0.20 to 0.50 SD, and are described by the MPR team as "more-favorable" or "relatively large." Under traditional demarcations, these effects would be considered "small" to "medium." In some cases small effects may be important, but in most circumstances we agree with the traditional demarcations. (See Appendix 1 for a further discussion of effect sizes and their interpretation.)

Sustained effects. The evaluation did not examine post-intervention impacts.

Benefit-cost analysis. Apparently not performed.

Cost-effectiveness analysis. Apparently not performed.

Commentary

John M. Love, Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet,
Jeanne Brooks-Gunn, Kimberly Boller, Diane Paulsell, Allison Sidle Fuligni,
Lisa J. Berlin, Jill Constantine, Cheri Vogel, and Christy Brady-Smith[1]

Douglas Besharov and his colleagues have graciously invited us to respond to their summary and critique of the findings from the Early Head Start national evaluation. Perhaps with this opportunity to clarify some misunderstandings about the findings and the study methodology, we can help other potential users of these important results see their significance for Early Head Start and other programs serving low-income families with infants and toddlers.

We first reported on the Early Head Start programs' interim impacts in the January 2001 *Summary Report*,[1] followed by the June 2001 technical report to Congress, *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Children in Low-Income Families,* which reported findings through the children's second birthday.[2] A year later (June 2002), we submitted the final evaluation report, *Making a Difference in the Lives of Infants Toddlers and Their Families: The Impacts of Early Head Start.*[3]

---

[1]John M. Love is senior fellow at Mathematica Policy Research, Inc., in Princeton, New Jersey; Ellen Eliason Kisker is senior researcher at Mathematica; Christine M. Ross and Peter Z. Schochet are senior economists at Mathematica; Jeanne Brooks-Gunn is Virginia and Leonard Marx Professor in Child Development and Education, Teachers College, Columbia University; Kimberly Boller is senior research psychologist at Mathematica; Diane Paulsell is a researcher at Mathematica; Allison Sidle Fuligni is in the Department of Psychology, University of California, Los Angeles; Lisa J. Berlin is with the Center for Child and Family Policy, Duke University; Jill Constantine is senior economist, Mathematica; Cheri Vogel is a researcher at Mathematica; and Christy Brady-Smith is with the National Center for Children and Families, Columbia University Teachers College.

[1]U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Commissioner's Office of Research and Evaluation and Head Start Bureau, *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families, Summary Report* (Washington, DC: U.S. Department of Health and Human Services, January 2001).

[2]John M. Love, Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Kimberly Boller, Diane Paulsell, Alison Sidle Fuligni, and Lisa S. Berlin, *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Technical Report* (Washington, DC: U.S. Department of Health and Human Services, June 2001).

[3]John Love, Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Diane Paulsell, Kimberly Boller, Jill Constantine, Cheri Vogel, Allison Sidle Fuligni, and Christy Brady-Smith, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start* (Washington, DC: U.S. Department of Health and Human Services, June 2002).

These reports, along with multiple volumes documenting the programs' implementation,[4] tell the story of the implementation and impacts of seventeen programs that were among the first grantees within a rapidly expanding, Congressionally mandated, national program. Importantly, these were not special demonstration projects but part of a large-scale, nationwide initiative. Beginning with sixty-eight programs authorized at 3 percent of the federal Head Start budget in 1995, the program now includes almost 700 grantees and is authorized at 10 percent of the Head Start budget. The seventeen programs, along with others, continue to operate within the national program beyond the evaluation period. The results of the randomized design yielded a large number of statistically significant impacts that formed a meaningful pattern of outcomes. Because none of the individual impacts were large, Besharov and his colleagues feel that we overstated their importance.

In this response, we explain why we do not believe the impacts are overstated and we show the weaknesses of their other critiques. Most important, however, we reiterate the main messages from the evaluation. In their focus on potential shortcomings in the evaluation, Besharov and his colleagues fail to note the most important policy message: that by analyzing subgroup differences, the Early Head Start evaluation has made important strides toward answering the major question challenging all program evaluations: "What works for whom under what conditions?"

What the Review Leaves Out

Program evaluations have often been criticized for providing overall impact findings while not learning about the conditions under which programs are more effective-often described as addressing the "What works for whom?" question.[5] The Early Head Start evaluation included a systematic assessment of program implementation, which enabled us to describe the programmatic approaches taken by the seventeen research sites as well as the degree to which they succeeded in implementing key requirements of the revised Head Start Program Performance Standards. Using subgroup analyses conducted within the framework of the experimental design, we compared impacts by program approaches, patterns of implementation, and child and family demographics.[6]

Supporting the overall impacts (averaged across all seventeen programs, all types of families, and all circumstances) showing favorable cognitive, language, and social-emotional

[4]For example, Administration for Children and Families. *Pathways to Quality and Full Implementation in Early Head Start* (Washington, DC: U.S. Department of Health and Human Services, December 2002).

[5]Administration on Children, Youth and Families. *Head Start Research and Evaluation: A Blueprint for the Future, Recommendations of the Advisory Panel for Head Start Evaluation Design Project* (Washington, DC: U.S. Department of Health and Human Services, September 1990).

[6]Note that although program characteristics were measured after random assignment, the assignment of families within programs did not change once random assignment was completed. Thus, Besharov, Germanis, and Higney are incorrect in asserting that the results of these subgroup analyses are not experimental findings.

outcomes for children and positive effects on parenting, parent self-sufficiency, and several aspects of fathering and father-child interactions, are several important stories from the Early Head Start evaluation: (1) reaching full implementation of the Head Start Program Performance Standards contributed to a stronger pattern of impacts; (2) although all program approaches had favorable impacts, the mixed-approach programs (those combining center- and home-based services in various ways) had a broader pattern of impacts; and (3) although patterns of impacts varied somewhat according to family characteristics, impacts were found for most types of families; of the twenty-seven subgroups studied, twenty-three experienced significant favorable child impacts and twenty-four had favorable parenting impacts (with stronger impacts for families who enrolled earlier, when the mother was still pregnant, for African American families, and for families with a moderate number of demographic risk factors); two difficult-to-serve groups also benefitted: parents at risk for depression and teenage parents, as well as such policy-relevant groups as welfare families, working families, and families headed by teenage mothers. We return to the important issue of subgroup findings later in this commentary.

How Criticisms Miss the Main Points

Questions about individual findings are used to discredit more widespread findings. The critique focuses on the "small" impacts found on the outcomes the evaluation measured. Besharov, Germanis, and Higney correctly suggest that researchers use the effect size metric as a way of comparing the size of impact estimates across different kinds of outcomes. We agree with this approach, and in fact presented the findings in both the original measurement metric (whether Bayley scores, rating scale points, or percentages) and in relation to the standard deviation of the measure (the effect size), as is often recommended (for example, Jacob Cohen, and Kathleen McCartney and Robert Rosenthal).[7] While various authors suggest criteria for labeling impacts of particular magnitudes "meaningful," in fact, these recommendations vary widely and, unfortunately, nothing in the effect size itself tells us the meaning of the program-control difference.

What researchers are able to do, however, is to compare the magnitude of effects across studies, and the critique quite appropriately takes this approach. This strategy, however, necessarily limits comparisons to outcomes that studies have in common. Thus, when the critique compares the cognitive impacts with those found in studies like Abecedarian, IHDP, and CCDP, it dismisses these admittedly "modest" (our term) impacts while isolating them from their larger context that our report described. Although some other studies have found larger impacts on particular outcomes, none of the studies that the critique compares Early Head Start to, or any other study that we are aware of, has found the wide range of significant impacts across multiple dimensions of children's development as well as multiple characteristics of the children's parents and families

[7]Jacob Cohen, "Things I Have Learned (So Far)," *American Psychologist* 45 (1990): 1304–1312; Kathleen McCartney and Robert Rosenthal, "Effect Size, Practical Importance, and Social Policy for Children," *Child Development* 71 (2000): 173-180.

that were found for Early Head Start. As both *Building Their Futures* and *Making a Difference* document, Early Head Start impacts included children's cognitive, language, and social-emotional development; parenting knowledge, discipline strategies, and the home environment; parents' mental health and family functioning; and parent self-sufficiency outcomes. Little is known about the potential cumulative effect of this large number of modest-sized impacts. Moreover, a number of measures on which we found impacts, and that are not highlighted in the review, are predictive of later functioning of infants and toddlers, e.g., children's social-emotional development, home environments, and parent-child interaction.

The review acknowledges only a small portion of the substantial subgroup impacts found. Other examples could be pointed out, such as (1) impacts on children's sustained attention in semistructured play and engagement of parent during play among children whose mother was pregnant at the time of enrollment (effect sizes of 0.47 and 0.52, respectively); (2) impacts on aggressive behavior problems (reducing them) and parent supportiveness during play among African American families (effect sizes of -0.35 and 0.47, respectively); (3) impacts on children's PPVT vocabulary scores, percentage of parents who read to their child daily, and parent participation in education and training programs among Hispanic families (effect sizes of 0.38, 0.29, and 0.30, respectively); (4) impacts on the percentage of parents who read to their child daily and on parent detachment during play (reducing it) among families with a moderate number of demographic risk factors (effect sizes of 0.33 and -0.34, respectively); (5) impacts on child sustained attention with objects and engagement of parent during semistructured play among children whose mothers were at risk of depression at the time of enrollment (effect sizes of 0.37 and 0.35, respectively); and (6) impacts on percentage of parents who read to their child daily, child engagement of parent during semistructured play, and percentage of children scoring below 85 on the PPVT-III among families in mixed-approach programs that were fully implemented early (effect sizes of 0.46, 0.43, and -0.34, respectively).[8] Although researchers may not be in full agreement on the issue, a number of writers believe the important information for policy about the effects of social programs are to be found in the subgroups rather than the global analyses, and this is particularly so when applying findings to continuous program improvement efforts.

Assumptions about unknown future results are used to discredit the known present ones. A theme running through the critique is that future events are likely to demonstrate the limited value of the Early Head Start impacts. Citing other evaluations, Besharov and his colleagues note that positive impacts found at one age are often no longer present at a later age. We argue, however, that a program that betters the present lives of families who live with a number of risk factors should be considered successful at the present time, regardless of what the future brings.

[8]See *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start, Volume I,* Chapter VII ("Differential Effects of Early Head Start Programs on Children and Families with Different Characteristics") and Chapter VI ("Variations in Impacts by Program Approach and Patterns of Implementing Key Features of the Performance Standards").

The critique uses a similar strategy in questioning the study's methodological rigor. When we reported that our analysis of attrition found that the impact estimates are not biased by differential attrition at this stage of the research, the study is critiqued on the grounds that attrition *could* become a problem in the future. We have no crystal ball that would permit us to counter this argument, but it is irrelevant for judging the worth of the present findings.

Furthermore, Besharov and his colleagues tell us that even if the program were found to have impacts at kindergarten, these would not guarantee continuing impacts into the school years. Again, this argument has no counter. The question for both policymakers and researchers is whether pessimism is the appropriate lens for guiding policies to support programs that have demonstrated short-term impacts. Our assessment attempts to be realistic rather than overly optimistic or pessimistic.

The absence of benefits with dollar values is used to discredit the potential economic benefits. Comprehensive, two-generation programs look expensive. Unfortunately, it is far too early in the lives of the Early Head Start children to be able to judge whether savings to society associated with the program impacts will offset these costs. Although the Early Head Start evaluation was not designed to be a cost-benefit study, a number of outcomes measured have potential cost benefits (such as reductions in the number of children who might require special services). Some impacts, on the other hand, such as increasing the percentage of children identified with a disability, will result in at least temporary increases in costs. Complicating this issue, any cost-benefit analysis would have to determine the costs of the extensive services that control group families received so as to calculate the net costs of the program. It is too soon to know whether Early Head Start will produce cost benefits.

Conclusions

In less than three years of providing services to low-income families with infants and toddlers, the seventeen Early Head Start research programs have produced impacts that appear promising in a number of respects. Unlike the samples found in most "model" or demonstration programs, the Early Head Start study sample is diverse and heterogeneous. Moreover, while many model or demonstration programs have not been implemented widely, or have difficulty demonstrating impacts when replicated on a large scale, the impacts of the Early Head Start intervention were found as the program was being implemented on a national scale, in typical communities across the country, as a regular part of the Head Start Bureau's program implementation strategies. Finally, it is noteworthy that the impacts were more numerous and larger in programs that were more successful in implementing the federally established and monitored Head Start Program Performance Standards. In order to disparage the evaluation's findings (and therefore the programs' accomplishments) and to consider the glass "half empty," one must assume, as Besharov and his colleagues do, that modest gains across a large number of dimensions cannot be cumulative, that these gains will disappear over time, that impacts within policy-relevant subgroups are not important, and that the program benefits do not justify the costs.

The research has been carefully done, and multiple approaches to the analyses attest to the robustness of the findings. Program operators, policymakers, and other researchers who want to know the details of the research methods and findings will find them in the technical reports we submitted in June 2001 and June 2002.

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/chapter7.html.