SCHOOL of PUBLIC POLICY

11
Even Start Family Literacy Program (3rd National Evaluation)

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/chapter11.html.

11
Even Start Family Literacy Program (3rd National Evaluation)

The Even Start Family Literacy Program was originally authorized by Congress in 1965 and reauthorized in 1994. The Even Start program is intended to: "help break the cycle of poverty and illiteracy by improving the educational opportunities of the Nation's low-income families by integrating early childhood education, adult literacy or adult basic education, and parenting education into a unified family literacy program."[1] In 2000–2001, there were 855 Even Start projects serving 31,896 families.

Robert St.Pierre, then vice president and principal associate at Abt Associates Inc., and his associates (the "Abt team") conducted the third national Even Start evaluation which covered the 1999-2000 and 2000-2001 program years. The evaluation used random assignment in eighteen sites and appears to have been implemented without problems. The research sample, however, was not nationally representative, having a considerably higher percentage of families that were Latino and that lived in urban areas. The program, which cost approximately $10,366 (in 2005 dollars) per family during the 2000–2001 program year,[2] apparently produced virtually no meaningful effects on a range of cognitive and behavioral outcomes. The Abt team suggests that one possible explanation for the absence of program impacts is that about one-third of the families in the program group did not participate in the program and many families in the control group received comparable services. A second possible explanation offered is the lack of any coherent curriculum content in the Even Start classrooms. A third possible explanation is that those Even Start families that did participate only participated in a fraction of the available services.

Program Design

Program group. At the time of this evaluation, Even Start was targeted to families with an adult who was eligible for adult education programs under the Adult Education Act or was

[1]Cited in Beth C. Gamse, Dylan Conger, Dean Elso, and Maria McCarthy, *Follow-Up Study of Families in the Even Start In-Depth Study: Final Report* (Cambridge, MA: Abt Associates Inc., 1997), 1.

[2]About half of this total was funded by the federal government and half by local contributions.

Assessments of Twenty-Six Early Childhood Evaluations

within the state's compulsory school attendance age and had a child younger than eight years old. During the project period, Even Start projects were required to recruit and serve families that were "most in need."

The families randomly assigned to Even Start were: 16 percent single-parent families; 16 percent with a high school diploma or GED; 75 percent Hispanic or Latino; and 27 percent employed.[3]

Services. Even Start grantees had considerable flexibility in designing services to meet the needs of the low-income families, but all were required to offer four services: (1) adult education to develop basic educational and literacy skills; (2) early childhood education services to provide developmentally appropriate services to help prepare children for school; (3) parenting education to help parents support the educational growth of their children; and (4) parent-child literacy activities. Even Start projects also offered referrals to a range of support services, such as child care, transportation assistance, health care, nutrition, and an array of other family support services. When these services were not locally available, some Even Start projects provided them directly.

The Evaluation. The third national evaluation of the Even Start program, funded by the Department of Education and conducted by the Abt team, included an "Experimental Design Study" (EDS) that consisted of eighteen projects.[4] To be included in the evaluation, the projects had to meet Even Start's legislative requirements, be operational during the 1999–2000 or 2000–2001 program years, offer instructional services of moderate to high intensity, and be able to serve at least twenty new families at the time random assignment was to take place. During the two years the Abt team recruited programs for the evaluation, only 115 out of about 750 programs met these criteria, and only 18 (about 16 percent of the 115 eligible projects) were willing to participate.[5] Eleven projects began participating in 1999–2000 and seven in 2000–2001.

About twenty-six families were randomly assigned in each site, with two thirds assigned to the program group and one-third to the control group. Because of the small number of families in each site, St.Pierre and his colleagues pooled the data for analytic purposes, resulting in a total of 309 Even Start families and 154 control group families. The analysis focused on all families,

---

[3]Robert St.Pierre, Anne Ricciuti, Fumiyo Tao, Cindy Creps, Janet Swartz, Wang Lee, Amanda Parsad, and Tracy Rimdzius, *Third National Even Start Evaluation: Program Impacts and Implications for Improvement* (Cambridge, MA: Abt Associates Inc. under contract to the U.S. Department of Education, Planning and Evaluation Service, 2003), 173, http://www.abtassociates.com/reports/2003107409351_80519.pdf (accessed July 2, 2010).

[4]The entire evaluation cost $6.7 million. The random assignment portion of the Even Start evaluation is a part of the total evaluation effort that cost $3.6 million.

[5]St. Pierre et al., 2003, 153.

regardless of the age of children, although the analysis of literacy impacts was restricted to children at least two-and-a-half years old.

Major Findings

For Even Start children and parents, there were few statistically significant effects on a range of cognitive and behavioral outcomes. The Abt team describe the findings as follows:

> Because we assessed the effectiveness of Even Start on 41 different outcome measures, we expected to see a few significant differences by chance alone. In fact, there are three significant differences between Even Start and control group participants. One of these favors Event Start and two favor the control group. Because of the large number of outcomes assessed and because of the mix in direction of results, we do not assign any meaning to these findings.[6]

Cognitive. There were no favorable statistically significant effects on parent reports of child literacy or a variety of cognitive tests, including the Peabody Picture Vocabulary Test (PPVT) and the Woodcock-Johnson Psycho Educational Battery (Revised) (WJ-R). Indeed, two of the subtests of the WJ-R indicated that Even Start children did worse than their counterparts in the control group.

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

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

Health. Data apparently either not collected or not reported.

Behavior. Even Start children were rated by their teachers using the Social Skills Rating System as having fewer behavior problems in elementary school (but not preschool).

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

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

Economic outcomes. There were no statistically significant effects on parent employment or annual income.

Effects on parents. There were no statistically significant effects on parent literacy, GED

[6]St.Pierre et al, 2003, 165.

attainment, or education level.

Benefit-cost findings. Apparently a benefit-cost analysis was not performed. The cost of Even Start during the 2000–2001 program year was about $10,366 (in 2005 dollars), of which 50 percent was provided by federal Even Start funds. Given the absence of positive findings, it is doubtful that the program produced savings to offset these costs.

Overall Assessment

The evaluation used random assignment in eighteen sites and appears to have been implemented without problems. The research sample, however, was not nationally representative, having a considerably higher percentage of families that were Latino and that lived in urban areas.

Program theory. Apparently, there is no specific theory detailed beside the general expectation that early intervention programs promote school readiness and improve developmental outcomes for children. The Abt team also hypothesizes that "Even Start also should have direct, but longer-term, effects on the economic self-sufficiency of adults due to increased participation in adult education and mediated by subsequent enhanced literacy skills."[7] Within this context, the evaluation is appropriate.

Program implementation. Program implementation was assessed by conducting two-day site visits, which included "interviews with staff from Even Start and collaborating agencies, observations of early childhood and adult education classes, and interviews about program costs." Seventeen of the eighteen projects provided participation data. In these projects, only 65 percent of the Even Start families participated in the program. High levels of nonparticipation, in turn, would be expected to limit the magnitude of program effects.

At the same time, many families in the control group received the same services through similar programs in the community, also possibly reducing the impact of the intervention. As the Abt team reports:

> Thus, the comparison made in the EDS is not between families that participated in Even Start and families that participated in no educational or social services whatsoever. Rather, the comparison is between families that enrolled in Even Start and families that participated in whatever mix of educational and social services that they obtained on their

[7]Robert St.Pierre, Beth Gamse, Judith Alamprese, Tracy Rimdzius, and Fumiyo Tao, *National Evaluation of the Even Start Family Literacy Program: Evidence from the Past and a Look Towards the Future* (Washington, DC: U.S. Department of Education, Office of the Under Secretary, 1998), http://www.ed.gov/pubs/EvenStart/index.html (accessed April 3, 2006).

own, in the absence of any assistance from Even Start.[8]

The projects spent 55 percent of the federal Even Start funds on instructional services, such as early childhood education, adult education and parenting education, and 9 percent on support services. The remaining 36 percent was spent on program administration, case management, and other activities. This pattern of spending was consistent with Even Start projects nationally, as well as over time.

Assessing the randomization. The Even Start projects in the evaluation were asked to recruit families "as they normally do" and then provide the list of families to the Abt team, who would then randomly assign two-thirds of the families to Even Start and one-third to the control group. Once assigned to the control group, a family could not participate in Even Start for one year. The total sample consisted of 463 families, with 309 in the program group and 154 in the control group. Thus, the ratio of program to control families was 2:1, as intended. The Abt team indicates that the families "were statistically equivalent at the time of randomization and at pretest." The groups were comparable on six of the seven background variables examined, although there was a statistically significant difference (at the 90 percent level) in the percent of single parent households (16 percent for Even Start vs. 23 percent for the control group).[9]

Assessing statistical controls in experimental and nonexperimental evaluations. Although the Even Start evaluation was designed as an experimental evaluation, some of the analyses included in the evaluation were not based on random assignment. For example, the Abt team suggests that "children who participated more intensively in early childhood education scored higher on standardized literacy measures."[10] This finding, however, is not an experimental finding and could be explained by other factors. The Abt team acknowledges: "Because amount of participation is a function of family characteristics (as well as program characteristics such as amount of service offered and the extent to which families are encouraged to participate) these relationships may also be explained by factors such as differences in the motivation of families or in their opportunity to participate in Even Start."[11]

Elsewhere, the Abt team indicates that nonparticipating families were included in the analysis, but "a separate set of analyses showed that omitting these families made no difference to

---

[8]St.Pierre at al., 2003, 163.

[9]The background characteristics examined included the percent of families: where Spanish is spoken at home; where English is spoken at home; that are Hispanic or Latino; with a high school diploma or GED; that are single parents; that are employed; and that have household income below $9,000 (when the enrolled in Even Start in 2000–2001).

[10]St.Pierre, et al., 2003, 152.

[11]St.Pierre, et al., 2003, 152.

the findings." This was an appropriate way to handle the issue. Simply omitting the nonparticipating families could severely undermine the random assignment process, since those who choose not to participate are likely to be different than those who do.

Sample size. The total sample consisted of 463 families, with 309 in the program group and 154 in the control group. The sample size for some outcomes was smaller (due to attrition) and some children were too young to be tested (under two-and-a-half years of age). The Abt team provided detailed information about the statistical power for some of the principal outcome measures. They explain that the evaluation "had very high statistical power to detect large and medium-sized effects, but poor power to detect small effects."[12] In accord with traditional demarcations, the Abt team defined large effect sizes as over 0.80 standard deviations (SD), medium as 0.50–0.8 SD, and small as 0.20–0.30 SD. (See Appendix 1 for a further discussion of effect sizes and their interpretation.) The Abt team argues that the sample size is large enough for policy-making purposes:

> We argue that while small effects may be interesting to researchers they are not always relevant for policy making purpose and, hence, that the statistical power offered by this evaluation is appropriate for determining the effectiveness of and improving Even Start. Even so, some may raise the question of whether the findings from the present evaluation would be seen in a different light if the EDS sample was substantially larger. If we assume that effects as small as 0.10 standard deviations were statistically significant for the EDS, then 17 of the 41 comparisons . . . between Even Start and the control group would be termed "significant." As many of these significant effects favor the control group as Even Start, so while a larger EDS sample might let us find additional significant differences between Even Start and the control group, we would have the same concerns about Even Start's effectiveness.[13]

Attrition. Between the time of randomization and pretest, 10 percent of the families were lost, but the researchers note that this attrition occurred evenly and did not affect the comparability of the program and control groups. The response rate for the pretest was 90 percent and 81 percent for the first posttest (about nine months later).

Data collection. The data collection relied on a wide range of standardized tests and surveys. The data sources were appropriate for the questions being studied.

Measurement issues. The Abt team used a broad range of measures to assess changes in literacy skills and various measures of school readiness, including direct assessments of the child,

[12]St.Pierre et al., 2003, 155.

[13]St.Pierre et al., 2003, 155.

parent reports of the child's skills, teacher reports of behavior, and school records. The Abt team used well-known tests to measure child outcomes and assessed the validity and reliability of those measures. Many other possible outcomes were ignored, probably because the intervention was not designed to affect them.

Generalizability. The eighteen sites with random assignment were not representative of Even Start projects generally. Projects were excluded if they did not meet Even Start's legislative requirements, did not provide high or moderate intensity services, or did not have the capacity to recruit thirty new families and conduct random assignment. In addition, the research sample differed from Even Start families nationally in a number of respects. For example, among families in the evaluation, 75 percent of the families in the research sample were Hispanic or Latino, compared with 46 percent of Even Start families nationally, and 83 percent of the research projects were in urban areas, compared with 55 percent of Even Start projects nationally. The Abt team contends, however, that the projects are representative of an important segment of the Even Start population:

> While such over-representation means that care should be taken in applying the findings to Even Start projects as a whole, almost 50 percent of the families served by Even Start are Hispanic and about 50 percent of the projects are in urban areas. Hence, the EDS findings do apply to an important and growing part of the Even Start population.[14]

Aside from differences in the characteristics of program participants, the Abt team cautions that the "projects volunteered for this study instead of being randomly selected, so we cannot generalize to the Even Start population on a strict statistical basis."[15] Moreover, the criteria used to select sites for the random assignment portion of the Even Start evaluation may have been geared toward showing positive results by restricting the projects to those that offered moderate to high intensity services.

Evaluator's description of findings. Despite the absence of program impacts, the Abt team may have obscured the program's failure by how they characterized the findings and by presenting a number of nonexperimental comparisons. For example, the key summary of the program's impact is stated as follows:

> While Even Start children and parents made gains on literacy assessments and other measures, children and parents in the 18 Even Start programs that participated in the EDS did not gain more than children and parents in the control group, about one-third of whom

[14]St.Pierre et al., 2003, 77.

[15]St.Pierre et al., 2003, 154.

received early childhood education or adult education services.[16]

The assessment begins by saying that "Even Start children and parents made gains," suggesting a positive impact, but later says that the gains did not exceed those of the children and parents in the control group. A clearer statement would have been to simply say there were no statistically significant effects.

Moreover, the researchers then use nonexperimental methods to examine the impact of the intensity of participation and various child and parent literacy outcomes. They found a positive relationship between the hours a child spent in early childhood education and various subtests of the RW-J. This, however, was a nonexperimental finding and subject to considerably more uncertainty than an experimental finding.

Overall, the Abt team concludes that, despite the lack of findings, the Even Start program should be strengthened.

> The fact that two experimental studies of Even Start show similar results, even though they were done at different times, one in the early 1990s at the very beginning of the program and a second after a decade of program implementation and many amendments to the program, points to the need to explore improvements if the Even Start model is to be an effective family literacy intervention. As implemented in the EDS projects, Even Start was not more effective than the mix of services that control group families obtained for themselves. Given Even Start's intuitive appeal as an approach for enhancing parent and child literacy, we interpret the lack of effectiveness as an indication that the Even Start approach needs to be strengthened.[17]

According to the Abt team, the primary program component in need of strengthening is the curriculum. The initial Even Start legislation focused on the implementation of the program, but did not address questions of curriculum content, let alone quality. For example, Even Start programs must first take advantage of related services that already exist in the community, before building their own curriculum and programs, which can be quite costly. In turn, the various components that make up an Even Start program differ at each site. As the Abt team explains, "It is difficult for an Even Start project to build a research-based program if it has not paid attention to, or has little control over, the quality of some or all the instruction received by its families."[18]

Replication. This is the second random assignment Even Start evaluation, also conducted

[16]St.Pierre et al., 2003, 3.

[17]St.Pierre et al., 2003, 10.

[18]St.Pierre, Riccuiti, and Rimdzius, 29.

by Abt Associates, but with a larger sample and stricter randomization procedures. Like the first evaluation, however, it too suggests that Even Start has few impacts.

Evaluator independence. Even Start was evaluated by an independent evaluator, Abt Associates.

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

Effect sizes. Effect sizes were calculated by dividing the difference in mean scores between the Even Start group and the control group and by the standard deviation (of the measure of interest) and were reported in standard deviation (SD) units.

For only one measure, the program group had a statistically significant gain with an effect size of about 0.35 SD. For two measures, the control group had statistically significant gains over the program group, with effect sizes of about 0.3 SD. The Abt team defined large effect sizes as greater than 0.80 standard deviations (SD), medium as 0.50–0.8 SD, and small as 0.20–0.50 SD. Thus, these effect sizes were considered small, which is in line with traditional demarcations. Additionally, as the Abt team notes:

> We argue that while small effects may be interesting to researchers they are not always relevant for policy making purposes, and hence that the statistical power offered by this evaluation is appropriate for determining the effectiveness of and improving Even Start. Even so, some may raise the question of whether the findings from the present evaluation would be seen in a different light if the EDS sample were substantially larger. If we assume that effects as small as 0.10 standard deviations were statistically significant for the EDS, then 17 of the 41 comparisons in Exhibit 6.12 between Even Start and the control group would be termed "significant". As many of these significant effects favor the control group as Even Start, so while a larger EDS sample might let us find additional significant differences between Even Start and the control group, we would have the same concerns about Even Start's effectiveness.[19]

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

Benefit-cost analysis. Apparently not performed.

Cost-effectiveness analysis. Apparently not performed.

[19]St.Pierre et al., 2003, 155.

Commentary

Robert G. St.Pierre*

The Department funded the third national Even Start evaluation in the fall of 1997. This study addressed questions about the outcomes and effects of Even Start projects, the kinds of families served by Even Start, the implementation of Even Start projects, and whether certain approaches produced better outcomes for Even Start participants. To address these questions, two complementary data collection efforts were undertaken: the Even Start Performance Information Reporting System (ESPIRS) and the Experimental Design Study (EDS).

The ESPIRS continued the decade-long annual collection of a common set of data from all Even Start projects. It provided information on the types of projects funded, the nature and amount of services they offered, the collaborative efforts they undertook, and the obstacles that existed to program implementation. The system also provided annual child, parent, and family level data, including demographic information, education and income data, the amount of service families received, and the progress they made on indicators of parent, child, and family well being, such as economic self sufficiency, literacy skills, and parent child relationships.

A second component of the third national evaluation, the Experimental Design Study (EDS), was included to provide a stronger assessment of program effects than had been undertaken in the past. The EDS used a research design in which families that wanted to take part in Even Start were randomly assigned to begin the program right away (intervention group) or to wait for one year (delayed intervention or control group). The design also included case studies of program operations and a study of program costs.

Projects were recruited during the 1999–2000 and 2000–2001 school years to participate in the evaluation. During this time, all Even Start projects in the nation were screened for eligibility. To pass the eligibility screen, projects had to meet Even Start's legislative requirements, be in operation for at least two years (ensuring that they had already dealt with start-up issues), plan to operate through the length of the study, plan to serve at least twenty new families at the start of data collection (necessary to meet the sample size requirements of the evaluation), offer instructional services of moderate or high intensity relative to all Even Start projects (there was no interest in studying projects that deliver low-intensity services since the Even Start legislation requires projects to deliver "high-intensity" services), and be willing to participate in an experimental evaluation of Even Start's effectiveness. Projects were recruited from urban and rural areas across the nation and served varying proportions of

*Robert G. St.Pierre was vice president and principal associate at Abt Associates Inc.

English-as-a-Second-Language (ESL) participants. Over two recruitment years, 115 out of a universe of about 750 projects met these selection criteria. All 115 eligible projects were contacted, materials were sent describing the study, telephone calls were made to all 115 projects to discuss the study, and site visits were made to more than 40 projects. In the end, eighteen of these projects (about 16 percent of those eligible) volunteered to participate in the study.

Each of the eighteen projects was asked to recruit families as they normally do and to be certain that all recruited families were eligible for Even Start, wanted to participate in Even Start, and were willing to participate in the evaluation whether they were assigned to Even Start or a control group. Each project provided listings of recruited families to research staff who randomly assigned those families either to participate in Even Start (two thirds of the families) or to be in a control group (one third of the families). Families that were randomly assigned to Even Start participated in the program at whatever levels of intensity and for whatever duration they desired. Families that were randomly assigned to the control group were not allowed to participate in Even Start for one year. During that year, however, they took part in any other educational and social programs which they sought out and for which they qualified. It would have been unethical to prohibit and impossible to prevent control group children/parents from enrolling in other preschool programs (e.g., Head Start), from enrolling in an adult education program, or from taking advantage of a locally available parenting education program. A total of 463 families were randomly assigned: 309 to Even Start and 154 to the control group.

After completion of the third national evaluation, the Department of Education secured a contractor to design and implement a large-scale study to improve the quality of instructional services in Even Start by identifying and testing the relative effectiveness of enhanced family literacy interventions. Information from this study should provide local family literacy projects with evidence-based strategies that they might adopt, instead of having to invent their own family literacy interventions.

In spite of these forward-looking steps, the fundamental hypothesis underlying the family literacy model—that the presence and integration of four instructional components will enhance literacy outcomes for children—remains unverified, and as of this writing the Bush Administration is proposing to eliminate funding for the Even Start program in the fiscal year 2006 budget.[1] Research points to the effectiveness of providing high-quality, center-based, early education

[1]Editors' note: Since the writing of this commentary, both the Bush and Obama administrations have proposed terminating the program, citing the three evaluations as justification. Federal funding for the program fell from $279 million in 2003 to $66 million in 2009. Gail McCallion, *Even Start: Funding Controversy* (Washington, D.C.: Congressional Research Service, January 17, 2006), http://www.ed.psu.edu/goodlinginstitute/pdf/CRS_even_start_funding_controversy.pdf (accessed July 2, 2010); and U.S. Department of Education, *Fiscal Year 2011 Budget Summary and Background Information* (Washington, DC: U.S. Department of Education, 2010), http://www2.ed.gov/about/overview/budget/budget11/summary/11summary.pdf (accessed April 8, 2010).

instruction to children, but as long as there remain persistent and troubling questions about our ability to enhance the literacy and parenting skills of adults, we cannot expect to see subsequent enhanced effects on children. Until we find a way to demonstrate that child literacy can be influenced by earlier positive effects on parent literacy and parenting skills, we need to continue to rely heavily on high-quality center-based early childhood education to help develop literacy skills in preschool-age children from low-income families.

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/chapter11.html.