

Executive Summary¹

Since its inception, the Even Start program has undergone evaluative scrutiny at a variety of levels. In addition to the ongoing federally-funded national evaluation, each local project is required by law to conduct an independent local evaluation, and in some states, projects participate in a state-wide evaluation. While the primary purpose of many state and local evaluations is to help improve and focus the program in specific locales, these evaluations may contain information that could be more widely disseminated to program developers and policy makers.

To understand what types of evaluation activities are being undertaken, and to see whether state and local evaluations do, indeed, contain information that would be useful for project improvement and helpful outside of their immediate applications, the U.S. Department of Education (hereafter, the Department) asked Abt Associates Inc. to collect, read, and report on state and local Even Start evaluations. This document is the final report from that project.

This project aimed to gather and review state and local Even Start evaluations, to describe the types of evaluations that were conducted, to describe findings about the outcomes of Even Start programs, and to provide guidance on improving state and local evaluations. In practice, the study focused on local evaluations since few state-level studies were available for review at the time that we were collecting evaluation documents.

Even Start Evaluation Requirements

Even Start was initiated as a demonstration program. After its first decade, it remains a program that is designed to help demonstrate and learn about the best ways of implementing and delivering family literacy services. To assist with this mission, strong evaluation requirements have been a prominent and continuing feature of the

1 This report was prepared by Robert St.Pierre, Anne Ricciuti, and Cynthia Creps at Abt Associates Inc. under purchase order 43-31KV-7-A2049 with the U.S. Department of Education. The views expressed in this report do not necessarily reflect the position or policy of the Department, and no official endorsement by the Department should be inferred. Staff at the Planning and Evaluation Service within the Department were responsible for providing technical input and oversight of the study. As Project Officer for this study, Tracy Rimdzius oversaw all planning, implementation, and reporting activities. Valuable comments on early drafts of this report were provided by Tracy Rimdzius, Valena Plisko, and Patricia McKee.

Even Start legislation. In addition to the ongoing national evaluation, the legislation calls for each project to “...provide for an independent evaluation of the program.” (P.L. 103-382, Sec. 1205). The only additional guidance on local evaluations comes from the Department:

“Each Even Start project is required to provide for an independent evaluation of the project. See section 1205 (10) of the Act. These evaluations provide local projects, States, the Department, and the Congress with objective data about the activities and services provided by the project, the participants served, the retention rates of those participants, and the success of the families in the project.” (Department of Education preliminary guidance).

Obtaining State and Local Evaluation Reports

During the 1996-97 school year there were about 650 local Even Start projects, and the law specifies that each project is to conduct a local evaluation. For the present study, we hoped to collect at least 100 local and state evaluation reports, though we had no idea how many reports could actually be obtained. Because some State Coordinators require their grantees to submit copies of local evaluations, we sent a letter to all State Coordinators, describing the planned synthesis and asking for their help in obtaining copies of state and local studies. Obtaining local evaluations was difficult. Having limited time and funds, we ended up with a sample of convenience – a total of 4 state evaluations and 118 local evaluations from 19 states.

Using Data for Program Improvement

Local evaluations are conducted for multiple purposes. First and foremost, Even Start projects have to comply with the legislative requirement for a local evaluation. Given this requirement, many local evaluations choose to focus on project outcomes in an attempt to provide evidence so that local project directors can obtain political and/or financial support for the program from school boards, civic organizations, corporate sponsors, or foundations.

The evidence contained in the reports that we reviewed shows that Even Start projects rarely engage in the systematic use of data to manage and improve their programs. Instead, program improvements/alterations typically are made on the basis of anecdotal evidence obtained through observations and stories gathered from the personal experiences of program implementers. A few reports that we reviewed noted recommendations from the previous year, described whether they had been addressed, and provided additional recommendations for the current year.

Local Evaluation Methods

Of the 122 evaluations reviewed for this study, almost all reported some information on the implementation of the projects being studied including descriptions of project structure and activities, level of participation, recruiting and retaining families, and working with collaborating agencies.

More than three-quarters of the evaluations we reviewed contained information about program outcomes. The most common design for local evaluations, used in 76% of the cases,² was to pretest and posttest Even Start families at the start and end of a school year. No control or comparison families were included in these studies. In 31% of the outcome studies Even Start families were not pretested, but they were posttested at the end of a school year. No control or comparison families were included in these studies. Only 10% of the local outcome evaluations used a quasi-experimental design in which the gains of Even Start families were compared to the gains of families in a non-equivalent comparison group, e.g., families in a parallel program, children in Head Start, etc. Finally, no local outcome evaluation used an experimental design in which families eligible for Even Start are randomly assigned to participate in the program or in a control group.

Evaluation Findings

We selected 24 of the 122 evaluations that focused on outcomes and read them carefully to understand their findings.

Outcomes for children. Local evaluations showed a consistent pattern of positive gains on several different measures of child development including the PreSchool Inventory (PSI), PreSchool Language Scale-3 (PLS), Peabody Picture Vocabulary Test (PPVT), pre-IDEA Proficiency Test, and the High/Scope Classroom Observation Record (COR). When significance testing was done, the gains generally were statistically significant. The size of the gains is consistent with the size of the gains observed in the national Even Start evaluation and in other family literacy studies which administered these same measures.

Outcomes for adults. As was the case for child-level outcomes, local evaluations showed a consistent pattern of positive gains on several different measures of adult development including the Comprehensive Adult Student

² Percentages add to more than 100 because some evaluations used more than one design.

Assessment System (CASAS), the IDEA Proficiency Test, and the Tests of Adult Basic Education (TABE). Statistically significant gains were observed in almost all projects that administered these measures. In addition, most projects that collected data on GED status reported that substantial numbers or percentages of adults attained a GED during their participation in Even Start.

Estimating program impacts. Half of the reports that we reviewed gauged the size of these outcomes by making comparisons to other studies. For example:

- ' ***Comparison to national data.*** Six studies compared local gains on the PSI, PLS, CASAS, and the Home Screening Questionnaire (HSQ) to average gains reported in the national Even Start evaluation. As would be expected, the gains in the 6 studies were larger than national gains in some cases and smaller in others.
- ' ***Comparison to national statistics.*** One project found that compared to data presented in the National Education Goals Reports, Even Start parents are more likely to read to their children than parents nationally.
- ' ***Comparison to gains observed in other evaluations.*** Some studies compared local Even Start gains to gains from other studies of adult education or early childhood education programs. For example, one project compared Even Start parents with parents in a local JOBS program on expectations of their child and view of parenting. No differences were found between the two groups of parents. Two projects found that the CASAS gains of their adults were greater than the gains of adults in California adult education programs. Yet another program compared the CASAS gains of their adults to those of adults in a Toyota-funded family literacy project. Finally, one program found that the PSI scores of Even Start children were lower than the scores of children in a Home Start evaluation conducted 20 years ago.
- ' ***Comparison to "normal" development.*** Some studies compared local Even Start gains to normal development based on the use of local or national norms. For example, one project found large gains when comparing developmental age with chronological age using PPVT scores; a second project projected normal growth based on pretest scores of the Learning Accomplishment Profile and found that Even Start children gained more than expected.

These local evaluations made an important attempt to judge the magnitude of the gains observed for Even Start children or adults. However, none of the studies was

able to estimate the “true” impact of Even Start, which would require randomly assigning incoming families to Even Start or to a control group. Given the resources available to local evaluators, it is unreasonable to expect that any local study could actually put a randomized experiment in place.

Recommendations

The Department currently spends much more on local Even Start evaluations (more than \$3 million per year) than on the national evaluation (\$500,000 to \$750,000 per year). If the mere existence of a local evaluation report is useful, or if it is useful for project directors to have a local evaluation that concludes that Even Start is a success, then local evaluations can be judged helpful to local projects. However, if the criterion for success is that a local evaluation should be used to help improve an Even Start project, then this review of more than 100 evaluations was not able to show that local studies provide much useful information, either to local project directors, to State Coordinators, or to the Department.

How can we improve this situation? Most important, in our view, is to provide clear guidance on the purpose of local evaluations. The current mix of guidance (to assess program outcomes, to address questions of local importance, to improve program functioning) needs to be streamlined and focused. We agree with the importance of the Department’s recent move towards a modification of the current system in which local grantees would be encouraged to use local evaluations as vehicles for program improvement. Some specific recommendations for improving local Even Start evaluations and for enhancing the ability of Even Start grantees to conduct continuous improvement efforts are given below.

Change the Even Start legislation and preliminary guidance to refer to “local continuous improvement efforts” instead of “local evaluations.”

Terminology is important, and the past and current use of the term “local evaluation” has become synonymous with a study only of program outcomes, with the implication that local projects will be in trouble if they cannot demonstrate positive outcomes. Use of the term “local continuous improvement effort” would more clearly signal the understanding that Even Start is a difficult program to implement, and the intention that projects are expected to engage in a systematic, ongoing assessment of the strengths and weaknesses, as well as the outcomes, of their approaches, and to use evaluation data to improve their programs.

Provide guidance to state coordinators and/or local projects on the amount of funding that needs to be spent on a good local evaluation which focuses on program improvement. Without such guidance new projects have no

idea of the amount of funds to allocate for this activity, and some do not include a local evaluation line item in their budgets. Suggesting a set amount or a percentage of the total project's budget that should be allocated to evaluation will signal the importance of evaluation and data collection to local program staff.

Help establish a community of local evaluators that can exchange information about useful evaluation approaches by making a mailing list of local evaluators available, by helping local evaluators exchange e-mail addresses, and by having sessions for local evaluators at annual Even Start conferences.

Provide training for local grantees in using data collected at the state and/or national level for continuous improvement at the local level. Local family literacy projects often regard participation in state-level and national-level studies as a burden that offers little or no return. State and national evaluation sponsors could help ensure local use of data by providing training to local staff in (1) accessing data collected for state or national studies, (2) understanding what those data mean at the local level, and (3) using those data to assess the performance level of their projects and to improve unsatisfactory performance.

Provide guidance to local projects on the use of data/evaluation for program improvement. Program improvement currently occurs through trial-and-error, and is based on the hard-learned experiences of program staff. This is one way of improving programs, but greater gains could be made by accompanying these personal, anecdotal methods with a data-based systematic assessment of program strengths and weaknesses. Such systematic continuous improvement efforts have been described by Haslam (1998) in a study of 12 Even Start projects, by Alamprese (1996) in studies of workplace literacy programs, and by Appel (1998) in her work with local Even Start evaluations.

The Department, in the Observational Study of 10 well-implemented Even Start projects, is facilitating and assessing the use of the continuous improvement approach based on the following principles:

Collaboratively set outcome goals for children and families. Local project staff, evaluators, and families enrolled in the program work together to set concrete outcome goals to be achieved.

Devise an intervention to achieve those goals. The family literacy intervention should be designed to achieve the specific outcome goals set in the first step. The intervention should be based on practices shown to be effective in prior research as well as on proven program quality indicators.

Set intervention thresholds necessary to achieve goals. An assumption underlying all family literacy programs is that if a family is to achieve their goals they

must participate in the program for a sufficiently long period of time at a sufficiently high level of intensity. Program staff need to define, ahead of time, the minimum intervention threshold that they believe is needed for families to achieve their goals.

Assess progress toward goals with sound measures. Assessing progress on a periodic basis by using sound measures (i.e., measures that have adequate reliability and validity, that have a history of use in similar studies, that are available in appropriate languages, and so on) is the best way of determining whether families have achieved their goals.

Use evaluation to monitor program quality and results, and to target areas for improvement. A comprehensive evaluation will help program staff monitor the quality of their family literacy program. It also will help staff keep track of each family's level and duration of participation in the intervention and assess progress towards goals. Knowing the quality of each program component, the extent to which families have participated, and the degree to which they are making progress will allow program staff to understand why some families do not achieve their goals (e.g., they did not participate at a sufficient level, or program components were of low quality), to identify program components that need improvement, and to target resources accordingly.