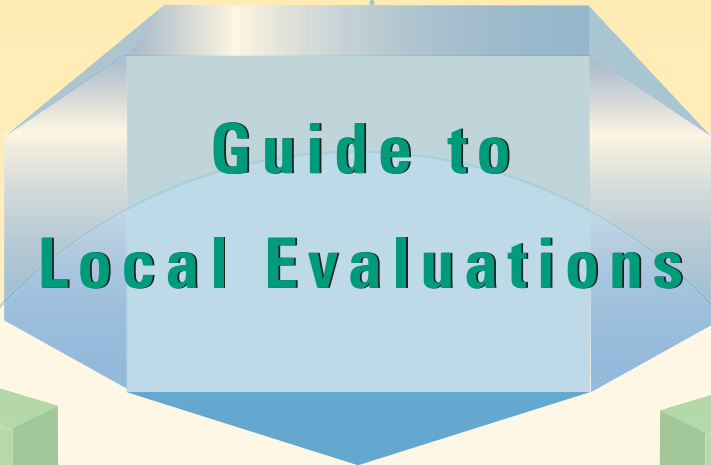


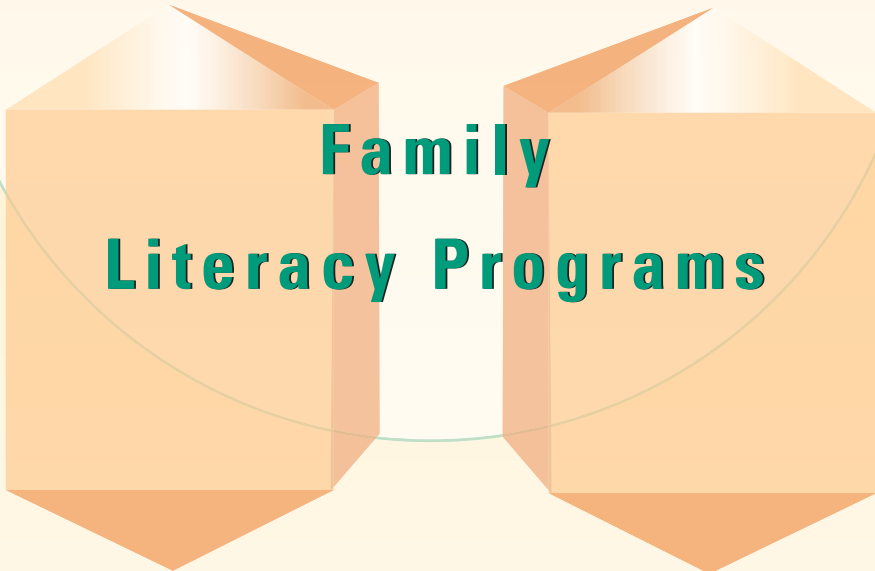
**DRAFT—Not for Duplication
Workshop Materials Only**



**Guide to
Local Evaluations**

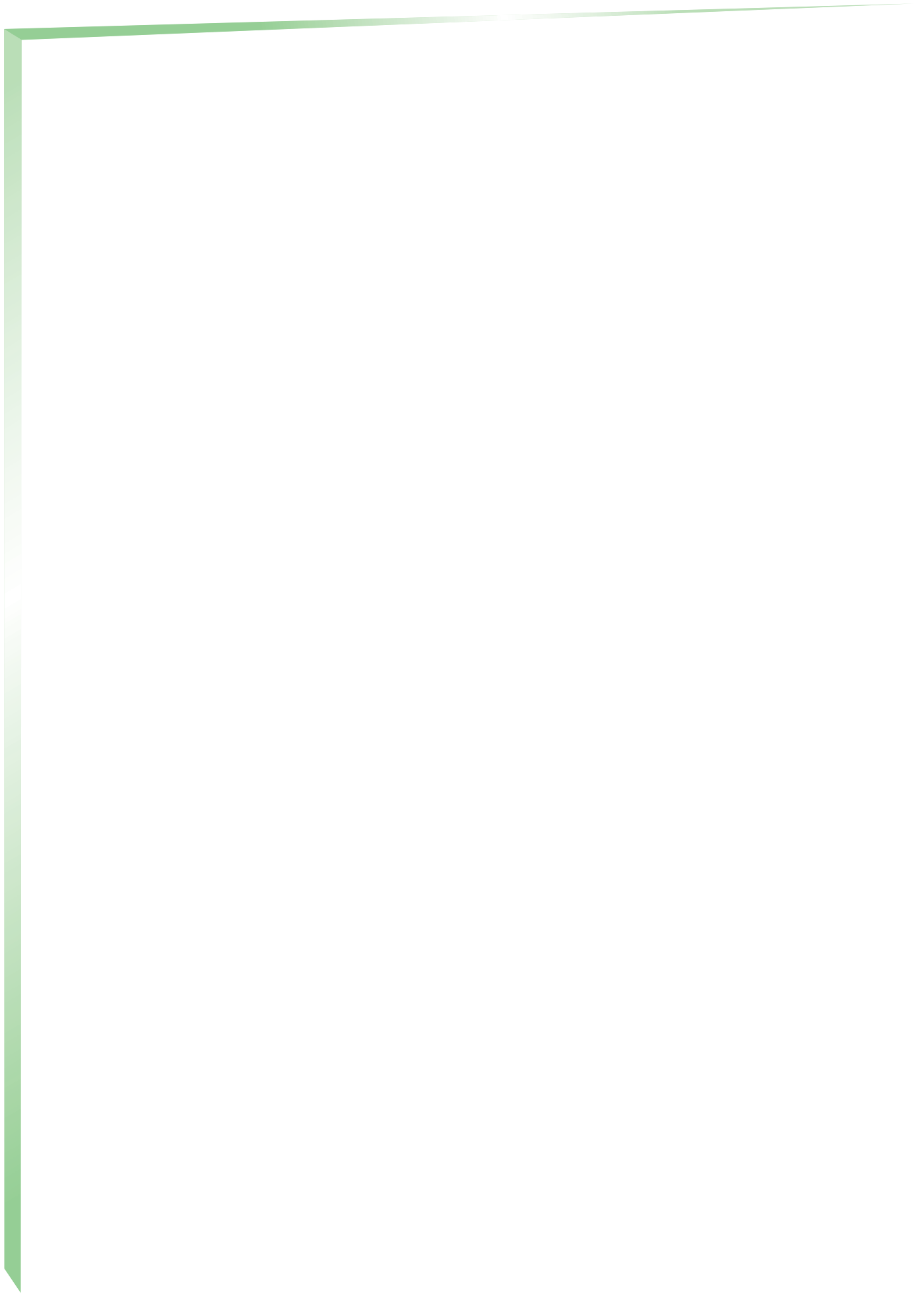
of

Even Start



**Family
Literacy Programs**

January 2005



Guide to Local Evaluations of Even Start Family Literacy Programs



U. S. Department of Education
Margaret Spellings
Secretary

Ray Simon
Assistant Secretary

January 2005

This publication was developed through a contract with the Center for Evaluation and Research (CER), ED-03-PO-2745, from the U. S. Department of Education, Student Achievement and School Accountability Programs, Even Start Family Literacy Program. The materials and resources mentioned, including assessment instruments or measures, program examples, and instructional techniques, are provided as resources and examples for the reader's convenience. The federal government does not endorse any particular book, article, author, assessment instrument or measure, publisher, or supplier, and inclusion of these references does not reflect their importance, nor is it intended to endorse any views expressed or information about particular assessment instruments or measures including statements about validity and reliability.

Writers: Eloise M. Appel of Appel Associates, M. Christine Dwyer and Susan L. Frankel of RMC Research Corporation, and Matthew B. Russell of CER

Designer: Diane Draper, RMC Research Corporation

Editor: Elizabeth Goldman, RMC Research Corporation

The document is in the public domain. Authorization to reproduce in whole or in part is granted. The citation should be: U.S. Department of Education, Guide to Local Evaluations of Even Start Family Literacy Programs by E. M. Appel, M. C. Dwyer, S. L. Frankel and M. B. Russell, Washington, D.C., 2005.

Copies of this document may be ordered in the following ways:

Mail: Write to ED Pubs, Education Publications Center,
P.O. Box 1398, Jessup, MD 20794-1398
Phone: 877-433-7827 or 800-872-5327
(TDD and TTY) 1-800-437-0833
Fax: 301-430-1244

E-Mail: edpubs@inet.ed.gov

Online: The document may be downloaded from the Department's Web site at: www.edpubs.org

On request, this publication is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at (202) 260-9895 or (202) 205-8113.

Table of Contents

Introduction	v
SECTION I:	
Chapter 1: Background	1
Key Evaluation Concepts	3
Lessons Learned	6
Federal and State Requirements	8
Chapter 2: Framework for Planning Evan Start Family Literacy Program Evaluations	11
The Conceptual Model for Planning Local Evaluations	12
A Tour of the Framework	13
Selecting a Focus	17
Role Distinctions Between Evaluators and Staff	20
SECTION II:	
Chapter 3: Participant Learning Outcomes	23
State-Required Indicators	24
Choosing Additional Outcomes at the Local Level	26
Selecting Instruments	27
Using Outcome Data for Continuous Improvement	30
Chapter 4: Refining the Program Design	35
Target Populations	37
Partnerships and Collaborations	40
Staffing and Facilities	42
Program Theory	43
Chapter 5: Program Content and Implementation	45
Curriculum and Instruction	46
Professional Development	48
Service Delivery	50
Quality Criteria for Program Self-Assessment	51
Chapter 6: Program Intensity and Participant Attendance	53
Key Evaluation Questions Related to Participation	54
Participation Variables	56
Managing and Analyzing Data	60
Using Participation Data	61
Chapter 7: Learning Milestones	63
Identifying Milestones	66
Identifying Measures to Assess Progress	68
Keeping Systematic Records	69
Reporting Progress	70
Troubleshooting Results	70
Relating Outcomes and Milestones	71
Using Milestone Data for Continuous Improvement	71



Table of Contents

SECTION III:

Chapter 8: Selecting an Independent Evaluator	73
Defining Independent Evaluation	74
Desired Qualifications of Even Start Evaluators	74
Setting Priorities for the Local Evaluation and Evaluator	77
Selecting an Independent Evaluator	79
Chapter 9: Working with an Evaluator	83
Basic Elements of an Evaluation Agreement	84
Developing a Scope of Work	84
The Evaluator–Staff Relationship	88
Chapter 10: Planning a Focused Inquiry	91
Choosing the Evaluation Focus	92
Developing Project Objectives and Evaluation Questions	94
Choosing Methods and Instruments	96
Developing a Timeline	97
Chapter 11: Reporting and Using Evaluation Findings	99
Maximizing the Use of Evaluation Findings	100
The Written Evaluation Report	101
Levels of Evaluation Reporting	102
Organization and Content of Reports	103
Enhancing Presentations of Evaluation Findings	106
Reporting Data Analyses	108
Chapter 12: The State Coordinator’s Role in Local Evaluations	113
Establishing and Communicating State Requirements and Policies	115
Providing Information, Training, and Support to Evaluators	116
Training Project Staff in Evaluation Concepts	116
Reviewing and Offering Feedback on Evaluation Plans and Reports	117
Using the Local Evaluation Results	118
Glossary	121

Introduction

The *Guide to Local Evaluations of Even Start Family Literacy Programs* is a resource for Even Start project directors, local evaluators of Even Start family literacy programs, and the state education agency (SEA) staff who oversee the Even Start family literacy program within each state. It is intended to provide common messages to those different audiences about the purpose and requirements of local evaluation in Even Start and the options to be considered in the design of local evaluations. Stronger local evaluation will improve the quality of Even Start programs and lead to better outcomes for Even Start families.

Each local Even Start project director selects an evaluator to plan and execute an evaluation that documents the achievement of participant outcomes and provides information useful for program improvement. Together the evaluator and project director determine the scope of the evaluation, frame evaluation questions, and identify data sources. Then, the task of the local evaluator is to provide an objective picture of participant outcomes and the factors that contribute to or inhibit progress. To be useful for program improvement, the evaluator must place the project's outcomes in context, and compare results and operations to state-established indicators and local project objectives. The local evaluator synthesizes a wide variety of quantitative and qualitative information to develop recommendations for improvement and then tracks the results of program changes over time.

The information in the *Guide* is designed to help readers think about the best choices for designing an objective evaluation that will have utility for the program. Evaluators and project staff can begin to consider the options for identifying evaluation questions and data sources by starting with the organizing framework presented in Chapter 2. Separate chapters (see Chapters 3 through 7) address each component of the framework (program design, program implementation, participation, learning milestones, and outcomes). Chapter 11 provides advice on organizing information from all data sources into a useful report.

Other chapters address frequently asked questions that have been posed by project directors and local evaluators. Chapter 1 addresses "big picture" issues about how evaluation in Even Start is different and similar to other types of evaluation and provides advice on requirements, scope, rigor, roles, and so forth. Chapters 8, 9, and 10 are about selecting and working with an evaluator, including developing agreements and plans. Chapter 12 addresses questions that state coordinators have raised about improving the quality of local evaluations.

The *Guide* is intended as a reference document to be consulted at key points in the evaluation process. Optimal uses for the *Guide* are:

- **Prior to selecting an evaluator**, project directors should begin by reading Chapters 1 and 2 to prepare for deciding on the scope and nature of the evaluation.

- Project directors will find Chapters 8 and 9 helpful in **selecting** and working with an evaluator.
- **Once an evaluator has been selected**, the project director should provide the evaluator with a copy of the *Guide*. Together they should discuss the material in Chapters 1 and 2 and then work through Chapter 10 to **develop the evaluation plan**.
- Depending on the direction of the evaluation plan, the evaluator can then consult Chapter 3 through 7 (each about one component of the plan) for ideas about **executing the evaluation**.
- **Once the evaluation is underway**, the project director and evaluator should read and discuss Chapter 11 to plan the **reporting of findings**.

The *Guide* does not include detailed information about the Even Start program requirements; a local evaluator will want to obtain information from the project staff or state coordinator about Even Start components and requirements. See information about program guidance at <http://www.ed.gov/policy/elsec/guid/evenstartguidance02.doc>. The *Guide* is not intended to be a substitute for expertise that can only be gained through formal training and experience in evaluation design and measurement. We recommend that programs hire evaluators who have appropriate qualifications and credentials (see more in Chapter 8).

A Word About Terminology

Terminology associated with roles varies by state so we have made some compromises. In this *Guide*, we have used “project director” to mean the staff member who has primary responsibility for the operation of the Even Start program and who is the person responsible for hiring and working with an evaluator. In some states, the role might be labeled project coordinator or co-coordinator. We have used the term “state coordinator” to mean the staff member who is designated by the SEA as the main point of contact for the Even Start program in the state and who would be charged with oversight of Even Start subgrantees.

Even Start subgrantees are partnerships between one or more LEAs (local education agency) and one or more nonprofit community-based organizations, public agencies other than an LEA, institutions of higher education (including two- and four-year institutions), or public or private nonprofit organizations other than an LEA. We use the term “partners” to refer to the agencies that comprise the Even Start subgrantee partnership. Other agencies outside the partnership that may provide services, e.g., transportation, health, early childhood education, are referred to as “collaborators.”

A Glossary at the end of the *Guide* defines technical and legal terms used in the text. Efforts were made to avoid using technical evaluation terms wherever possible, given the diverse audience intended for the *Guide*.



Chapter 1

Background

This chapter focuses on the foundations for local evaluations in Even Start— what’s expected of evaluations and why. It links local evaluation to program improvement. Chapter I covers:

- key evaluation concepts;
- lessons learned; and
- federal and state requirements

Even Start family literacy is a complex program designed to provide high quality and intensive literacy services for families with great educational and economic needs. Success requires well-qualified staff members who can pinpoint literacy needs, provide instruction based on scientific research, and motivate families’ commitment to sustained participation. Full participation also depends on community partnerships that provide the range of supports that families need in order to address literacy goals.

Because the needs of families vary, the Even Start program is necessarily flexible. Depending on a local population’s needs and partners’ resources, Even Start family literacy programs vary by configuration, staffing, and instructional level. Local populations vary enormously—single teen mothers with infants and toddlers, new immigrants from extended families who do not speak English, adults with learning disabilities who lack the skills for employment, incarcerated mothers with young children, and others.

The Even Start law focuses on the family unit with individual attention to a range of family members—children from birth through age seven and their parents with limited literacy skills.

Policymakers have challenged family literacy providers to design and provide services that are sufficiently rigorous and extensive to produce meaningful differences in family members’ literacy outcomes. *The local evaluation statutory requirement is part of addressing those challenges.* The Even Start statute includes local evaluation as one of fifteen required program elements. The designated purpose of local evaluation is program improvement.

Requiring an independent local evaluation is somewhat unusual for a program of such small scale. (Even Start programs typically serve about 30 to 50 families.) The requirement for independent local evaluations is intended to provide family literacy programs with data they can use to develop and improve high quality, intensive services that can make a meaningful difference in literacy outcomes for families in widely varying circumstances.

***“SEC. 1235. PROGRAM ELEMENTS.
Each program assisted under this subpart shall —
15) provide for an independent evaluation of the program, to be used for program improvement.”***

Independent means that the evaluator is not a member of the program implementation staff: program leaders are expected to benefit from an objective, evidence-informed perspective on whether or not families are making progress and

on the factors contributing to or hindering progress. The very complexity of Even Start—four core instructional components, support services when necessary, a wide range of children by age, high needs families—may make it more difficult for project staff to gauge meaningful progress. An independent perspective can help determine appropriate benchmarks, distinguish between services that might be supportive but not sufficient for making real progress, sort out strategies that are and are not working, and make recommendations for change.

Evaluators provide evidence-based advice. Evidence of progress toward outcomes is crucially important and motivating for staff working with hard-to-reach families. Evidence that shows lack of progress and points to redirection is especially critical in settings where staff may feel that caring attitudes alone are enough. The ultimate clients for the evaluator’s work are the families in Even Start programs. The results of the independent evaluator’s work improve the chances that families will reach important educational goals.

In many if not most cases, this ambitious vision for the role of local evaluation has not been realized. In 2000, Abt Associates prepared a report, *Synthesis of Local and State Even Start Evaluations*, which revealed that evaluators rarely included recommendations in their reports and that projects rarely used data systematically to manage and improve programs and results. Project directors with little evaluation experience may not have the expertise to steer evaluation planning in a useful direction or may find it difficult to identify qualified evaluators who understand family literacy. Resources for local evaluation have often been meager. Too often, Even Start evaluations have been simply summaries of existing descriptive data already collected by programs, or non-systematic observations of program operations without discussions of outcomes. Neither approach provides guidance for the continuous improvement of Even Start programs as envisioned by law.

Key Evaluation Concepts

Throughout this document are key assumptions about evaluation in Even Start. Summarized below, these assumptions could serve as the opening for a planning discussion between project directors and evaluators, and for training in local evaluation concepts.

- **Purpose of local evaluations.** Local Even Start evaluations are primarily designed for program improvement, that is, to identify what is and is not working well and the reasons why, in order to make changes that will improve outcomes for families. Evaluation is about learning—learning how to assess progress in attaining results and learning how to make a program more effective in achieving results. To facilitate learning, evaluations must yield information that project staff can use.

Even Start supports integrated family literacy services for high needs parents and children, primarily from birth through age seven, and has three interrelated goals:

- *to help parents improve their literacy or basic educational skills;*
- *to help parents become full partners in educating their children; and*
- *to assist children in reaching their full potential as learners.*

The term “family literacy services” is defined in section 9101(20) of the Elementary and Secondary Education Act of 1965 (ESEA) as services provided to participants on a voluntary basis that are of sufficient intensity in terms of hours, and of sufficient duration, to make sustainable changes in a family, and that integrate all of the following instructional activities:

- (A) Interactive literacy activities between parents and their children.*
- (B) Training for parents regarding how to be the primary teacher for their children and full partners in the education of their children.*
- (C) Parent literacy training that leads to economic self-sufficiency.*
- (D) An age-appropriate education to prepare children for success in school and life experiences.*

From The William F. Goodling Even Start Family Literacy Programs Part B, SUBPART 3 of Title I of The Elementary and Secondary Education Act of 1965 (ESEA), amended by the No Child Left Behind Act of 2001 (NCLB). Unless otherwise noted, all further section references in this document are to the ESEA.

- **Program theory.** Evaluation planning begins with a shared understanding by the evaluator and program staff of the program’s theory—how particular activities and approaches are intended to produce short and long term results for participants. A good grasp of program theory and design is foundational to an evaluator’s ability to recommend meaningful program improvements.
- **Central role of outcomes.** All local Even Start evaluations should report on and discuss participant learning outcomes, including specifically the results on the state’s performance indicators and local objectives for participants. Each state collects information within a set of common topics and reports the data annually to be aggregated at the federal level according to the GPRA (Government Performance and Results Act) indicators (<http://www.ed.gov/about/reports/annual/2004plan/edlite-evenstart.html>).
- **Analytic vs. advocacy perspective.** Local evaluations should focus on exploring areas that need improvement in order to recommend changes. Simply reporting information about program characteristics and participant outcomes is a part of evaluation, but is inadequate for improving programs. Evaluations may document what projects do well but their primary purpose is not to “rubber stamp” project decisions or to be public relations tools.
- **Phase of development.** The focus of evaluation will vary by a project’s stage of development; evaluation approaches need to match the local project’s stage of development and reflect its needs. In first-year projects, for example, more emphasis should be placed on success in implementing the program design, including setting up systems for capturing baseline information about participants on entering and throughout their participation in Even Start. Evaluations of projects that have been operating for several years may take a longitudinal look at the trajectory of participants’ progress and track the effects on participants after they complete the program.
- **Process and outcomes.** Evaluation for program improvement implies attention to program operations and processes as well as participant outcomes. The evaluation of even a first-year project should measure participants’ progress toward outcomes as well as program implementation in order to determine which aspects of the program are contributing to or hindering participant progress.
- **Stakeholder involvement.** Key staff, collaborators, and advisors together formulate the important questions that the evaluation should answer and also interpret and use the results. An independent evaluator is a key collaborator in this team effort and not a sole actor.
- **Evidence base.** Local evaluations are based on the systematic collection of information from many sources; they provide objective evidence about project outcomes and operations. Observations by the evaluator and recommendations for improvement must be based on evidence collected by the evaluator, the staff, or from program records. It is not sufficient for the evaluator to assume that the project is operating in exactly the manner described in the proposal nor is it adequate to gather information solely from the project director.
- **Qualitative and quantitative methods.** Local Even Start evaluators draw from both qualitative and quantitative methodologies when planning comparisons, selecting instruments and data collection procedures, choosing approaches to analysis, and presenting findings. For example, a statistical analysis of a

series of adult progress tests might be augmented by qualitative systematic observations of adult education instructors and interviews with adult students.

- **Systematic and objective approaches.** A corollary to the above statement is the expectation that evaluators will employ systematic and objective approaches. Objectivity rests on the nature of data, the way data are collected, and the way information is summarized and interpreted. Even Start evaluations frequently use interviews, focus groups, and observations to gather perspectives from staff and parents and to understand the quality of program services. Those methods are useful if executed systematically, that is, if questions are carefully designed to address relevant variables (specified in advance) and responses are treated as data in an analytic manner. Similarly, objectivity is greatly enhanced by gathering information from multiple sources about the same phenomenon, e.g., reviewing home visit records, observing home visits, and interviewing home visitors as well as parent participants. Ideally, the evaluator will use state performance indicators and other data collected through the course of program activities in several ways. For example, information about attendance patterns summarized by site opens a window into program operations; the same data might also be used to compare the performance of those who have been enrolled in programs for different lengths of time.
- **Standards and expectations.** Embedded in the concept of evaluation is a determination of value, that is, comparing findings to expectations—whether the comparison is with the performance of a reference group, a level of expected mastery, or established standards of quality. Simply describing what has happened in a project, e.g., discussing observations and interviews or test results, is not enough for evaluation. Instead, the evaluator should make relevant comparisons: comparing the observations to standards of quality for instruction, examining how interview information depicts what was expected, and/or placing test results in the context of normative development, of students in other programs, or of individuals' previous rates of change. Statewide results for performance indicators are a key source of comparison for local projects.
- **Audiences.** The primary audiences for Even Start local evaluations are project directors, staff members, and advisory boards associated with the local program. Secondary audiences are state Even Start coordinators and those charged with technical assistance to help local programs improve. Because an evaluation report has numerous audiences, it should be a “stand alone” document with enough context that readers from different perspectives can understand it.
- **Evaluation vs. research.** Even Start evaluations measure local program implementation and outcomes for family participants. They are not primarily designed to generalize results beyond the local program. In other words, local Even Start programs are not conducting research to build field knowledge. For that reason and given the scope of Even Start evaluations, it is not realistic to expect that local evaluations would include all the elements of research designs (see more about rigor in Lesson #3, below).

Lessons Learned

We recommend approaches in this *Guide* that take into account the practical realities faced by Even Start evaluators. Challenges that program directors and evaluators face include: the scope of resources for local evaluation, identifying the right mix of expertise and experience, the inappropriateness of traditional design approaches, the limitations of measures included in state performance indicators, the difficulty of measurement in the family literacy field, and the relationship between program staff and evaluators. Below are ways to address these challenges.

Lesson #1. Evaluations require adequate resources and a specific focus. State coordinators and local program directors have not always set aside sufficient resources to plan and execute a useful local evaluation. Annual Even Start local evaluation budgets top out at \$10,000-\$15,000, the range appropriate for carrying out the type of evaluation described in this *Guide*. Smaller budgets mean that project staff and evaluators cannot analyze outcomes and also target a known area of concern for additional data collection, e.g., *to what extent do early childhood staff consistently implement language production strategies and literary curriculum as intended?* Where evaluation resources are limited, it may not be possible to carry out such a focused inquiry, meaning that the evaluation falls short of what is expected of Even Start programs.

In some cases, limitations on resources might mean that evaluators do not collect specific data but simply rely on information the program already collects to meet state evaluation requirements. We recommend that at minimum, evaluators identify a specific focus for new data collection each year and also analyze outcome data collected for performance indicators. It is essential that programs allocate reasonable resources to evaluation (see more in Chapters 8 and 9).

Lesson #2. There is no substitute for evaluation expertise. In trying to maximize resources, program directors sometimes choose persons for the role of evaluator who have no training in evaluation but may provide other needed expertise to the program, e.g., an early childhood expert who can provide curriculum guidance. When programs employ evaluators without training in evaluation, they sacrifice both systematic data collection and a trained evaluator's objective eye.

Even Start evaluators should have had training in evaluation design, measurement, and data analysis as well as background or experience in one or more content areas relevant to Even Start. Thus, for example, a university professor in the early childhood department who also has had some training in research and measurement might be the perfect evaluator choice. Evaluators are assumed to have the expertise to apply general evaluation concepts (see Chapter 2) to specific applications, e.g., to summarize different types of assessment scores appropriately or to set up a database to track program information. The *Guide* does not reproduce the type of information found in evaluation texts but shows how and where to apply evaluation procedures in the Even Start context.

Lesson # 3. Project scale affects the rigor of design. When planning evaluation designs, local evaluators may be challenged by some features of Even Start, including the small program populations, the continuous enrollment policies, the great variation in intensity of treatment within the program population, and the selection criteria related to population definition (i.e., those with the greatest needs for literacy services). These features complicate traditional quantitative designs that depend on comparisons between groups (randomly assigned or matched in some other way) and measurements at fixed intervals, and have challenged even large-scale evaluations of Even Start programs.

Local Even Start evaluators often need to rely on comparisons based on norms, whether from a test's norming population or performance results from other programs, e.g., from a state's performance indicator reporting. Given the small sample size of most Even Start programs, local evaluators must think creatively about comparisons and analyses: combining results over several years to look for patterns, employing designs that use multiple measures for individuals so that individuals can be compared to their own baseline performance, pooling data with other programs with similar populations, and making various types of internal comparisons. For example, Even Start evaluators may look carefully at dosage (amount of instructional contact time) as a modifying variable (a way to sort categories) since individuals at the same site and performance level may vary greatly in their cumulative and current program year contact hours of instruction. (See more in Chapter 6.)

Thus while Even Start evaluation designs are not intended to provide generalizable findings, local evaluators should use the most rigorous designs possible, and thoughtfully plan the selection of measures, data collection schedules, and appropriate comparisons.

Lesson #4. Encourage projects to expand beyond state performance indicators. Evaluators help identify appropriate outcome and progress measures for program and participant goals. Although some states do not require common measures for program evaluation, the advent of state-level performance indicators (aggregated at the federal level via GPRA indicators) has advanced measurement in Even Start. In some cases, state-identified measures may not be a good match for specific contexts and populations. In those cases, evaluators help program staff select valid outcome measures that will augment state-prescribed measures. For example, the state-prescribed early childhood measure may focus only on reading readiness literacy skills measured at the beginning of kindergarten, but a local program's child population may be largely below the age of three. In that case, the evaluator would help identify appropriate language measures of child outcomes that link with school readiness outcomes.

Evaluators should also help review and identify measures and procedures that are used to track progress toward outcomes. (See more in Chapter 7 about learning milestones.) Periodic progress monitoring can help explain why outcomes have or have not been achieved. For example, a review of school-age children's records of skills mastered during supplemental tutoring sessions might shed light on why some children have not met end-of-year expectations, that is, which skill deficiencies remain.

Evaluators should have input into the instruments used to measure local progress and outcomes and incorporate the results of those measures in their evaluation analyses.

Lesson #5. The search for appropriate assessments involves compromise. Instruments that assess early childhood learning (including tests for non-English speakers), adult learning (especially English as a Second Language), parent-child interactions, and parent education may not be as well developed as instruments in many other fields. These learning areas pose special measurement issues in terms of validity (content and predictive) as well as reliability. In order to gauge the range of behaviors of interest reliably, stronger instruments tend to require considerable administrative time and expertise. Group testing may not be appropriate. Further, staff members are often frustrated by the futile search for a “perfect” instrument which simply may not exist. Evaluators can help in many ways: working through the various trade-offs posed by different instruments during the selection process, assisting with administration issues and scoring of assessments, and guiding interpretation. Evaluators should collaborate with staff in identifying and selecting instruments for measuring participant progress and outcomes.

Lesson #6. Effective evaluators engage with the program. While the local evaluator takes an independent perspective, his or her analysis is based on evidence that has not been gathered in isolation from program operations. It is important at the outset that program directors and evaluators work closely to identify the focus of evaluation, including areas of concern or program weakness that may need further study. Once data have been analyzed, the evaluator can help program staff interpret the meaning of evidence and shape recommendations for improvement that are feasible in the context of the local program. To do so, the evaluator needs a working knowledge of the Even Start program gained through review of materials, observation of instructional components, and discussion with staff and families. Given the range of populations, contexts, and approaches to program delivery, it is not advisable to conduct an Even Start evaluation “on paper” or at a distance from program operations. Evaluators and Even Start staff should establish a working relationship informed by familiarity with project operations.

Federal and State Requirements

Specific requirements for the required Even Start local evaluations may vary by state, and program directors and evaluators must be familiar with any state expectations. However, the essential roles and purposes of the independent local evaluation are the same for all projects, that is, to provide critical information for continuous program improvement to lead to better participant outcomes.

Differences in state requirements have typically emerged in response to the lack of specificity and utility of local evaluations. Some state policies define evaluator roles and provide basic requirements for evaluation reporting. A few states have prescribed evaluation requirements in detail, including specific tasks and information to be collected by all evaluators.

Some states prescribe approaches for one aspect of local evaluation; for example, states commonly require that evaluators include results for state performance indicators in the annual evaluation report. Simply complying with that state requirement is not sufficient, however, to fulfill the federal Even Start statute's requirements for a local evaluation that provides information which can be used for program improvement. Evaluators need to go beyond mere reporting to discuss ways to enhance outcomes based on data collected. This *Guide* offers project directors and evaluators ideas for integrating the basic federal and state requirements.

This *Guide* is designed to complement specific state expectations by providing an overall framework within which those state requirements fit. No matter how open or prescribed the state policy about evaluation, however, all Even Start local evaluations should address certain minimum expectations:

- Local evaluators summarize, analyze, report, and discuss data about participant outcomes (the state-required performance indicators as well as locally-determined group and individual outcomes and benchmarks) and explain results, including possible reasons for lack of progress;
- Local evaluators collect data about local project objectives;
- Local evaluators develop focused evaluation questions about program areas that need to be improved and focus data collection on those areas;
- Local evaluators collect evidence objectively and synthesize that information; and
- Local evaluators provide specific recommendations for program improvement based on evidence.

Local evaluators and project directors should take into account the results of the state's required indicators as well as objectives identified in the local project's approved application when identifying the focus for local evaluations. Evaluators can help programs set up systems to collect state-required data accurately, analyze that data, and base recommendations on those findings along with other information collected for the local evaluation.

Chapter 2

Framework for Planning Even Start Family Literacy Program Evaluations

Do **families** participate intensely enough and stay long enough to achieve goals?

How much **progress** have adults and children made toward intended outcomes?

Focus Questions

Are the **service providers** the correct mix to match families' needs and goals?

Are **instructional** services of high quality?

Examples for Evaluation Focus

Year 1

- *The evaluation plan for a first-year program checks on the degree to which the proposed program design has been implemented with quality, focuses on start-up issues, and ensures that systems are in place to collect baseline data.*

Years 2-3

- *The evaluation plan for a program with a dynamic enrollment analyzes the literacy outcomes of adults and children according to the extent and duration of their participation.*

Year 4

- *The local evaluation plan for a mature program aimed at improving its early childhood component augments the analysis of performance outcomes for adults and children with an in-depth look at the trajectory of children's progress on readiness milestones.*



This chapter introduces the framework that organizes the discussion of evaluation planning in the remainder of the *Guide*. Chapter 2 covers:

- the conceptual framework for planning local evaluations;
- a tour of the framework;
- selecting a focus; and
- role distinctions between evaluators and staff.

In Chapter 1, we described local Even Start evaluations as including, at a minimum, accurate and representative information about participant outcomes and local project objectives, exploration of an area of program concern (sometimes referred to as a focused inquiry), objective collection and synthesis of information, and recommendations for program improvement. There are many ways to design evaluations to meet these expectations and many ways evaluators can frame evaluation questions and combine data collection methods to learn about program processes and outcomes.

Because each of the four core instructional components of the Even Start program model has potential for in-depth exploration, the range of options for crafting an evaluation plan can overwhelm evaluators and program staff. While it is important to be aware of and consider various options before finalizing evaluation questions and firming up an evaluation plan, creating an effective and feasible evaluation plan involves zeroing in on one or two areas of concern.

The Conceptual Model for Planning Local Evaluations

The conceptual framework (see the figure on page 14) is a way to organize the discussion of evaluation planning and consider the range of options.

The framework's five "building blocks" show the logic of Even Start—how strategies are intended to produce outcomes. At the same time, the building blocks also represent potential sources of evaluation questions. The framework puts family literacy program theory together with the systematic thinking of evaluators.

In a given year, only some combinations of building blocks will be targeted for in-depth review and new data collection (although evaluators also often work with data that project staff members have previously collected). For example, an evaluator tracking children's progress on various milestones is likely to focus on collecting data from benchmark measures administered over time. She may not plan to collect new data about early childhood instruction systematically, e.g. via classroom observations, but she will need some background information from staff about the curricular goals that inform instruction in order to interpret the results of progress monitoring.

Although Even Start evaluators may not always collect new data or formally analyze information about each block in a given year's evaluation (other than participant outcomes), to some degree each framework building block plays a part in any evaluation plan.

Evaluation plans typically call for collecting data to answer specific questions. By addressing different questions each year, the evaluator can provide in-depth information on all program components and areas over the course of a four-year cycle.

The framework organizes the next five chapters. Chapters describe the evaluation questions, topics, and strategies that make up each building block: participant learning outcomes (Chapter 3); program design (Chapter 4); program implementation (Chapter 5); participation (Chapter 6); and learning milestones (Chapter 7).

A Tour of the Framework

The first feature to notice about the framework on page 14 is that it does not follow the typical linear form of logic models and conceptual frameworks.

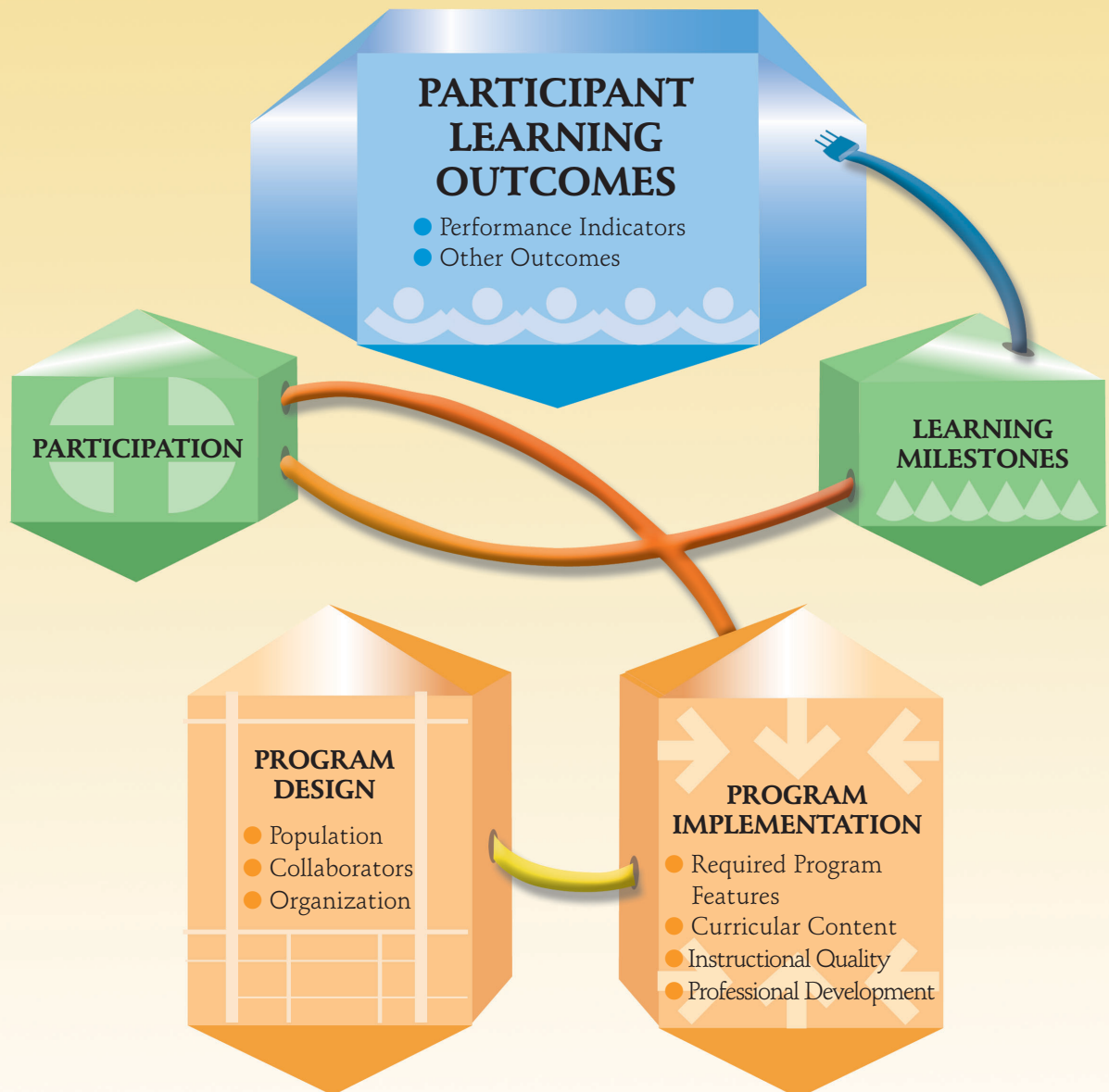
A “through line” (the thick cord) runs through the building blocks—from program design to program implementation to participation to learning milestones to participant learning outcomes—to show the path of relationships among the building blocks. Of course, different combinations of building blocks may interact—the assessment of the quality of program implementation might stimulate re-consideration of some aspects of program design, or less-than-ideal participant learning outcomes could spark examination of participation patterns.

Objective information about any one building block will affect the interpretation of data about other areas and stimulate new questions. The framework shows that the cycle of evaluation planning is continuous in the same way that the cycle of program improvement is continuous.

The second important feature of the framework is the prominent placement of **PARTICIPANT LEARNING OUTCOMES** at the top. No matter what other areas evaluators examine, all local Even Start evaluators should review and report on participant learning outcomes each year, disaggregating results to find patterns. Evaluators should present the results for the state's performance indicators along with other important learning outcomes, including any project objectives included in the approved application. Basic questions about outcomes always include:

- *Are projects meeting state-required performance indicators?*
- *Are projects meeting the outcome objectives stated in their approved applications?*
- *Are adults and children achieving other important learning outcomes?*

Framework for Planning Local Evaluations



The “thick cord” shows the most common path of relationships among the building blocks of evaluation, that is, design affects implementation which affects outcomes—mediated by the extent of participation by families and their achievement of learning milestones. See examples on pages 20 and 21 for other paths of interactions.

Evaluators also use participant learning outcomes for other purposes. Outcomes are the backdrop for understanding the influence and effectiveness of program strategies, and for examining the effects of varying degrees of participation on progress. (See more about outcomes in Chapter 3.) The two areas in the bottom layer of the framework, **PROGRAM DESIGN** and **PROGRAM IMPLEMENTATION**, form the foundation for examining program processes and quality. Even though basic features of program design, e.g. the population targeted, the key partners and so forth, are specified at the time of grant award, wise evaluators work with project leaders to examine the viability of those features over time. Note: The basic scope and objectives of the approved project cannot be changed without consultation with the state coordinator.

Common questions about design include:

- *How have changes in community demographics affected recruitment?*
- *Does the basic program model continue to meet the needs of families recruited for the program? If not, what types of refinements are needed?*
- *Are the service providers the correct mix to meet families' needs?*

Problems with participation or insubstantial outcomes may trigger an examination of program design at any point in the project life cycle. Evaluation tasks that can inform design changes include assessing needs systematically, determining alignment between resources and needs, and setting priorities. (See more about evaluating program design features in Chapter 4.)

Evaluators make major contributions to Even Start program quality by collecting data on the delivery of instructional services. Needing data to inform and justify modest refinements as well as more substantive changes, program leaders value objective perspectives on program operations.

Common questions about implementation include:

- *Is the program implemented as intended?*
- *Are instructional services of high quality?*

Many Even Start families face challenging circumstances that may affect their motivation to participate consistently in Even Start. Staff need to be well-prepared for and supported in their roles in order to motivate families. Evaluation questions about staff and family perceptions can also be important in evaluating implementation:

- *Do staff members feel prepared for their roles?*
- *What are participants' perceptions of the value of program services?*

Problems with participation and disappointing levels of progress frequently prompt a detailed evaluative review of the quality of program delivery, especially the delivery of instructional services. Methods for assessing implementation are wide-ranging, and include surveys and record reviews as well as interviews and observations. As described in Chapter 5,

standards for comparison are as important for interpreting the results of implementation evaluations as they are for understanding the magnitude of participant gains on learning measures.

The framework on page 14 shows **PARTICIPATION** and **LEARNING MILESTONES** in the “middle layer” between program-related areas and outcomes. This placement is intended to convey the role of these two areas as intermediaries for understanding program outcomes.

Much attention has been paid in Even Start to the intensity and duration of **PARTICIPATION**. The combination of high expectations and a high needs population leads, as one would expect, to wide variations in patterns of participation. Similarly, the flexibility of the Even Start model to adapt service delivery mechanisms to populations means that service intensity also varies widely. As the framework shows, participation can influence or mediate results in other areas.

Common questions about participation include:

- *Do programs offer sufficient hours of service and an adequately flexible schedule for each age cohort to realize intended outcomes?*
- *Do participants attend sufficient hours of service to realize intended outcomes?*
- *Do participants stay in programs long enough to achieve their goals?*

As Chapter 6 describes, program staff members may need help setting up systems to collect participation data accurately. Once systems are in place, the evaluator can use participation data to answer questions about process and implementation and also analyze data to understand other results. Examples of the latter are comparing the results of progress measures or performance indicators for participants with high and low rates of participation, or categorizing sites according to rates of family retention and participation and examining the differences in practice by site to see what factors help motivate participation.

LEARNING MILESTONES has a prominent position in the framework because focusing evaluation attention on intermediate steps has potential for improving programs. Evaluators can help project staff set up systems to track the progress of adults and children toward outcomes and help staff members understand the resulting patterns for the most common type of evaluation question:

- *How much progress are participants making toward achieving short-term learning outcomes?*

While progress toward benchmarks can be helpful in understanding why outcomes have or have not been achieved, the results can also be used to pinpoint areas of need within program operations and instructional processes for further examination. Finally, as described in Chapter 7, tracking progress on benchmarks motivates staff members and participants—encouraging continuous improvement on the part of staff and parents.

An evaluator will consider all five of the framework's building blocks—participant learning outcomes, program design, program implementation, participation, and learning milestones along the way to outcomes—and their interrelationships when crafting an evaluation plan. The evaluation plan will always address participant learning outcomes. While any one of the other building blocks might be a target for focused inquiry and new data collection, the evaluation plan will likely emphasize one or two areas. The choice of focus depends on the stage of program development, issues or concerns that have been identified, and resources available for local evaluation.

Selecting A Focus

Even Start projects are typically funded in four-year cycles. Areas of focus and evaluation questions change as the project evolves, and evaluation plans should look different in each year of the cycle (and in subsequent cycles). As project staff members become clearer about how to maximize outcomes for families, new strategies emerge to be tested. Expectations for what can be learned from local evaluations also increase over time.

During the **start-up year**, Even Start staff members often struggle to put all the required program elements in place. In the first year, project leaders are hiring qualified staff, ironing out partnership agreements, choosing curricula and securing instructional materials, recruiting families, putting accountability systems in place, and so forth. As a result, some first-year projects are slow to enroll families and get them participating regularly in all four components. Program energy is directed toward implementing the proposed project design.

Evaluation for a first-year project should focus on program design and implementation. Appropriate evaluation questions for the first year might be:

- *Are families participating consistently in all four core instructional components?*
- *To what extent are curriculum objectives, instructional materials, and pedagogical approaches research-based?*
- *Which recruitment strategies seem most effective for targeting the population in greatest need of literacy services?*
- *How well do program partners and collaborators understand the goals of Even Start, and do they agree with the goals?*

During a program's first year of operation, the evaluation should include baseline assessments to measure participant learning outcomes even if participants have not yet been in the program long enough to qualify for inclusion in the state's performance indicator reporting. The evaluator works with first-year program staff to ensure understanding of the required outcomes and measures, identify additional learning outcomes and measures appropriate for the population served, set up systems to track and record outcomes, and perform quality control checks on the initial administration of assessments and recording of data.

It may take a while for staff members to become skilled in administering and scoring instruments, especially instruments to assess language. The first year offers a good opportunity to provide staff with training and guided practice and to work out any misinterpretations associated with test administration such as setting basal and ceiling items, timing assessments, clarifying or re-stating directions vs. coaching, and so forth. This work—especially setting up data collection forms and data management systems—lays the foundation for accurate reporting in future years. All Even Start projects must put basic record keeping systems in place to respond to state-required and local information needs.

By the **second year of operation**, program staff members often struggle to retain families long enough to make important gains and try to address some families' erratic or limited participation patterns. In addition to tracking learning outcomes, the evaluator may focus on family participation with questions such as:

- *What are the participation patterns of Even Start adults and children?*
- *How is participation related to progress on benchmarks? (learning milestones)*
- *Which program features encourage participation? (program implementation)*
- *What are the barriers to participation? (program implementation)*
- *How do low and high participators differ in their success with outcomes? (participant outcomes)*

The range of questions above implies attention to more than participation alone, of course. Depending on the scope of the local evaluation, the evaluation plan might involve data collection in one or more areas.

The evaluations of **mature programs** are likely to reflect the sophistication of refined strategies, attempts to maximize outcomes, an understanding of different approaches for families with different profiles, and/or interest in rigorous documentation of outcomes. Evaluations of mature programs might employ a variety of ways to assess program outcomes, including measures that supplement required state performance indicator measures, analyzing the relationships among the results of different measures, assessment of longer-term effects (following up families that have successfully exited Even Start), and/or various subgroup analyses.

In the preceding examples, the program's stage of development suggested major issues and concerns related to program success. However, there may be times when mature programs face issues that newer programs experience. For example, recruiting and retaining eligible families may emerge periodically as a challenge even for long-standing programs. Certainly, lack of learning gains can be a problem for programs at any stage of development—it is especially notable as a problem given the small enrollments in Even Start programs.

Evaluators and project directors should expend evaluation resources on the most serious issues facing programs, that is, those issues that are hindering participant progress. This means that evaluations will often focus on program implementation and participation, and raise questions about the quality and intensity of instruction such as:

- *How do the oral language development opportunities available to Even Start children vary by classroom?*
- *To what extent do home visit plans reflect areas of educational need identified by the diagnostic assessments?*
- *How much more instructional time are Even Start children receiving compared to others of the same age?*
- *How much instructional time is required for adults who begin at different levels to achieve a grade level gain in reading and mathematics?*

In most cases, project directors and staff members are aware of concerns that warrant further exploration, and evaluators can elicit those concerns in initial planning discussions. In some cases, evaluators may need to prompt discussion about other sources of concern, such as comparing a program's results with statewide averages or raising questions about areas that monitoring visits have identified as weak. Components or whole programs may be officially designated by the state as low performing as a result of monitoring visits or during annual program progress reviews. Project directors should share any and all concerns with the evaluator to assure that evaluation plans will concentrate on the areas that most need improvement.

Projects facing several serious issues may need to allocate more resources to evaluation to address several areas and gather continuous feedback about strategies. For example, a floundering project may need the evaluator's help to monitor monthly participation, observe and provide feedback on the implementation of instructional emphases in all component areas, design and select progress monitoring tools, and assess participant motivation.

This multi-year example is provided to illustrate how formal inquiries build upon annual outcomes. The focused inquiries carried out by Rosemary Hodges, the evaluator of the Eastern Wayne County (NY) Even Start Program, in addition to annual reporting of outcomes asked:

Year 1

- *To what extent do Even Start parents understand and use recommended language strategies with preschool children?*

Year 2

- *To what extent is Even Start's relationship with local primary schools perceived as supportive? Does Even Start's support make a difference for children who are not meeting the school's grade level benchmarks in reading?*

Year 3

- *How does the support of a literacy specialist result in changes in the knowledge and teaching behaviors of family educators with preschoolers? Has the literacy specialist made a difference for children who are not meeting the school's grade level benchmarks?*

Year 4

- *How does the support of a literacy specialist result in changes in the knowledge and interactions of family educators with infants and toddlers?*

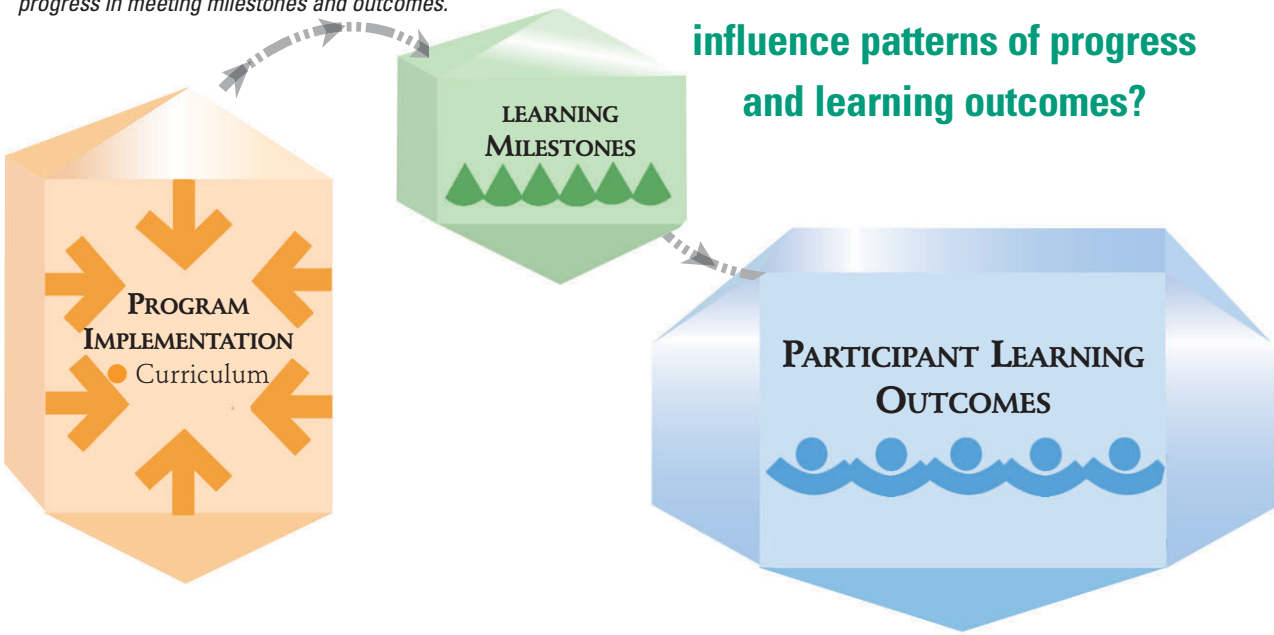
Role Distinctions Between Evaluators and Staff

Project directors, staff members, and evaluators play complementary roles in planning, executing, and using the results of evaluations. In a sense, the evaluator is a special type of project collaborator—a member of the team who brings an independent perspective to project concerns.

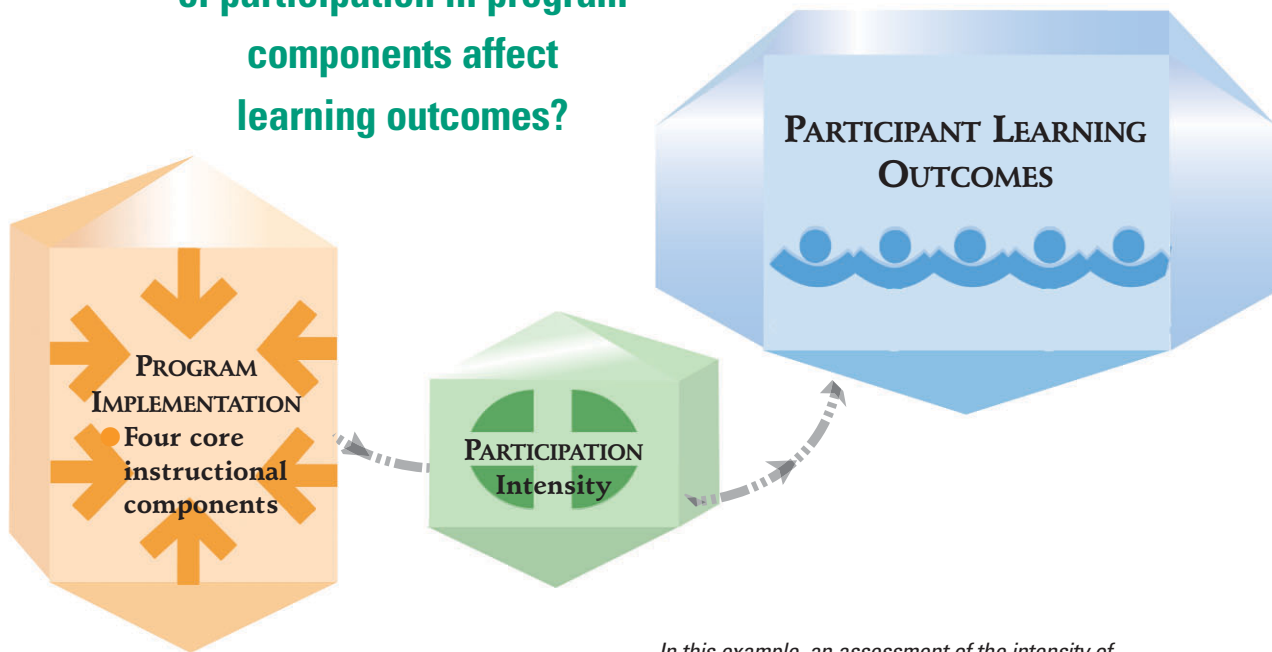
Understanding the boundaries between staff and evaluator roles helps to maximize resource use and ensure good working relationships. Essentially, a project director hires an evaluator to exercise professional judgment in collecting data about program issues and analyzing the resulting information. In practice, the evaluator and project staff members may together refine the questions and approaches for data collection. The project director has the ultimate responsibility for identifying the information needed to resolve local issues and for entering an agreement with a skilled and experienced evaluator who can provide evidence that will inform decisions.

Evaluators often use data that staff members have collected as part of ongoing project routines. Staff members may require special training to implement data collection procedures in an accurate, systematic, and fair manner. Involving staff members in recording data increases the likelihood that they will find evaluation results credible and useful. When staff members assume at least some of the responsibility for data collection, the project can use the evaluator’s time more wisely for analysis and interpretative tasks.

In this example, the assessment of program implementation (different curricula used at different sites) stimulates a look at participants’ progress in meeting milestones and outcomes.



How do varying rates of participation in program components affect learning outcomes?



In this example, an assessment of the intensity of participation stimulates examination of how family participation affects learning outcomes.

The next chapter begins a series of chapters about each of the five building blocks in the framework. **PARTICIPANT LEARNING OUTCOMES** is placed first because the measured achievements of adults and children are central to Even Start evaluation. When families have achieved important outcomes, evaluators can document how projects realized those successes. When families have not achieved or made meaningful progress toward outcomes, evaluators can help unpack the reasons by analyzing evidence collected through focused inquiries.

In these ways, looking at project services through the lens of outcomes helps keep services on track. Further, policymakers must rely on outcome information collected with the help of local evaluators, some of which is aggregated at the state level and then provided to the federal Even Start office via the GPRA indicators, to determine the value of Even Start.



Chapter 3

Participant

Learning Outcomes

The statute mandates states to develop indicators for adults in:

- *Achievement in reading, writing, English language acquisition, problem-solving, and numeracy;*
- *Entry into postsecondary education, job retraining program, employment or career advancement, including the military; and*
- *Receipt of a secondary school diploma or GED.*

And for children in:

- *Improvement in ability to read on grade level or reading readiness;*
- *School attendance; and*
- *Grade retention and promotion.*



Chapter 3 focuses on participant outcomes—the achievements expected of all adults and children in any Even Start project. It describes how to measure and analyze outcomes in terms of the state-required indicators and how to develop additional project-identified outcomes. This chapter covers:

- state-required indicators;
- choosing additional outcomes at the local level;
- selecting instruments; and
- using outcome data for continuous improvement.

State-Required Indicators

The Even Start law broadly identifies the important learning outcomes expected of adults and children enrolled in Even Start projects. Although Congress defined the parameters, it requires each state to develop its own performance indicators.¹ Therefore, indicators vary by state. State-developed indicators identify educational achievements expected of adults and children enrolled in local projects. States report on these indicators every year in their consolidated reports, and some of these data can be aggregated at the federal level with the GPRA indicators.

Some examples of indicator statements follow. Parentheses contain elements that should be included, such as the specific subject under study (e.g., adults who have completed 100 hours of instruction), the measure (e.g., CASAS), the performance or behavior (e.g., increased score 3 points), etc.

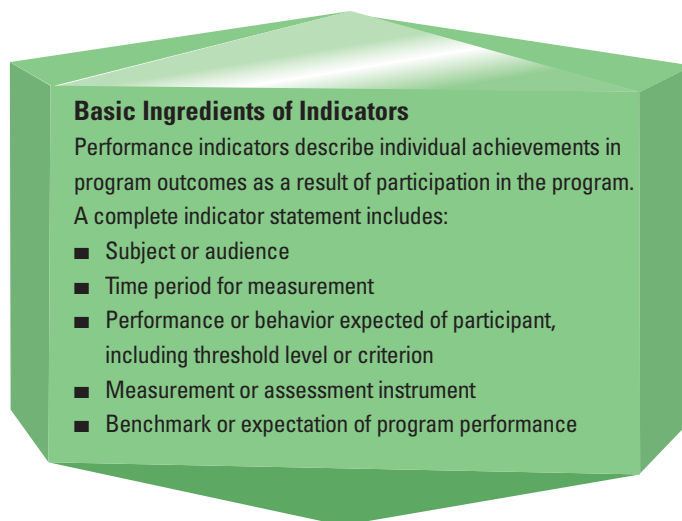
- **For preschool-age children:** 60% (program expectation) of preschool children (subject) will increase a minimum of one percentile ranking on the auditory comprehension or expressive language subscales (behavior or performance) as measured by the Preschool Language Scales (PLS) (measure) after eight months of participation (time period).
- **For adult learners:** 60% (program expectation) of adult learners who participate in adult literacy services for at least 100 hours (subject) will demonstrate an increase of four or more points over their entry-level test scores (performance or behavior) on each CASAS subtest (measure).
- **For English Language Learners:** 50% (program expectation) of all parents who pre-tested at levels 1, 2, or 3 on the BEST and who have completed at least 100 hours of instruction for English Language Learners (subject) will demonstrate a one level gain (behavior or performance) as measured by the BEST (measure).

¹ Throughout the *Guide* we refer to the state-required indicators as “performance indicators.” The statute calls them “indicators of program quality.” Many states use different terms for required indicators, such as “performance standards,” “state literacy indicators,” “quality indicators,” etc.

- **For school-age children:** 90% (program expectation) of third grade students in families that have reached enrollment status (subject) will be promoted to the next grade level (behavior or performance) as reported by the End of School Year Progress Report (measure).

All of the required Even Start indicators measure learning. The intent of the statute was for each state to develop a set of measurable learning outcomes that reflect high-quality family literacy practices. Indicators also often align with state standards for adult and K-12 education. For example, several states used the state attendance rate to set expectations for school attendance. The challenge for each state is to develop a set of indicators that fit its context.

In addition to the areas mandated by Congress, state coordinators may develop indicators for other outcomes. For example, states commonly add readiness indicators for 3-5 year olds, adult learning outcomes for English language learners, and parenting education outcomes. Where possible, these indicators are tied to existing state standards. For example, preschool indicators are usually tied to state-level early childhood standards.



Further examples of additional outcome indicator statements:

- **Parent-child interactive literacy activities:** 60% (expectation) of parents in families that have reached enrolled status (subject) will read to or look at books with their children three to four times a week (performance or behavior), as reported by pre-test and post-test parent-child literacy activities (measure).
- **Parents' support for children's learning:** Within in a program year, 80% (expectation) of parents who participated a minimum of 55 hours (subject) will visit the local library regularly to borrow books or other reading materials (performance or behavior) as measured by a parent-child family literacy rating scale (measure).
- **Adult English language learners:** 50% (expectation) of adults who have completed 75 hours of adult education instruction within the program year and pre-tested at Levels 0-3 on the English As A Second Language Oral Assessment (ESOLA) (subject) will demonstrate an increase of one grade level (performance or behavior) as measured by the ESOLA post-test (measure).

State indicators vary in their levels of detail and uniformity, especially in assessment instruments. For example, some state-developed indicators of literacy development for 3-5 year olds specify the instrument, others offer a choice of instruments, and still others provide no or vague guidance.

Role of project director. Project directors contribute their knowledge of the local context and provide access to local information. For example, if the evaluation requires access to school records, the project director will most likely have or establish contact with school staff.

Project staff members also have an overall sense of how instruments can be applied in their program. Project staff and evaluators usually decide on the instruments together.

Role of evaluator. Evaluators work with staff members to develop effective and efficient systems for collecting indicator data. The evaluator will likely spend a lot of time during a new project's first year setting up data collection procedures that are compatible with staff procedures. Over time, the effort of developing data collection procedures typically diminishes, and staff members will probably pay more attention to identifying local evaluation information that supplements and enriches the data required by the state.

Another important evaluator role is analyzing and interpreting state indicator and other project data so that results are meaningful and useful for project staff. Often this analysis involves disaggregating the data into meaningful subgroups, such as age, hours of participation, etc. (See page 30 for further discussion on using outcome indicator data to demonstrate program effectiveness and guide program improvement.)

Choosing Additional Outcomes at the Local Level

Local Even Start project directors and evaluators may also go beyond state-required indicators to measure project-specific outcomes. Given the variety of populations and contexts of local projects, general state-required performance indicators may not be appropriate for some groups served by the project, such as teen parents, migrant families, etc. Project directors may also want to add indicators to enhance the measurement of state-required learning outcomes. The project objectives listed in the application are a good source of additional participant outcomes to measure. For example, for school age children, outcomes beyond attendance and promotion could include grade changes in Math and English Language Arts.

Some examples of additional project-level education outcomes include:

- **For reading readiness:** Even Start kindergarten children will respond correctly to 50% or more of the items on the Concepts About Print test.
- **For reading:** 75% of children served by Even Start will continue to read at or above grade level in grades 3 and 4.
- **Gains in parenting literacy skills:** 60% of parents who score at Levels 1 (has knowledge) and 2 (able to understand) on the Family Literacy Parenting Scale will advance to the next level on the post-test.

In choosing outcomes it is important to consider:

- Is the outcome one the program can influence in a significant way?
- Will measuring the outcome help staff identify aspects of program implementation that can be changed?
- Do staff and other community stakeholders view the outcomes as important?

Role of project director. Staff members are instrumental in knowing and articulating the program aim. Although the local evaluator can facilitate staff discussion, the director and staff know the context and details of their program. The director also knows the kinds of outcomes that are most important to the local population. For example, if a project has a large number of English language learners, outcomes related to English language acquisition will be most meaningful to the program and the community.

Role of evaluator. Evaluators help program staff identify and develop outcomes that reflect the program's purpose. Through discussion, evaluators and staff determine measurable expectations of participants after they have engaged with the program over a specific time period, such as a year.

Selecting Instruments

As noted, some indicators specify the assessment instrument; in other cases, the project may choose among instruments. In this case, the evaluator can highlight each instrument's strengths and weaknesses and help staff members choose the best instrument, given the project's context (such as participant characteristics and instructional goals) for measuring an outcome. It is unlikely that any one instrument will be ideal.

When indicators call for performance data, such as a General Equivalency Diploma (GED) or school attendance rates, the evaluator may simply design a reporting mechanism rather than an instrument.

Evaluators need to consider several factors when selecting and using instruments to measure outcomes. Perhaps most important, evaluators should choose an instrument that permits individual outcomes to be aggregated, or combined, so that project staff and the evaluator can discuss the program as a whole. Another consideration is the appropriateness of the instrument for both the pre-test and post-test because a typical approach for measuring change is to look at differences between pre- and post-test scores.

Norm-referenced instruments. Norm-referenced instruments measure knowledge and skills in reference to a comparison group. This means that scores can be compared to a norm, or reference, group according to certain characteristics, such as age, grade-level peers, etc. Norm-referenced tests are helpful for measuring preschool outcomes because they provide a comparison for tracking learning development.

Examples of norm-referenced assessment instruments:

Children:

- *Peabody Picture Vocabulary Test 3rd Edition (PPVT-III)*
- *Preschool Language Scales (PLS-4)*
- *Test of Early Language Development 3rd Edition (TELD-3)*
- *Test of Early Reading Ability 3rd Edition (TERA-3)*

Adult:

- *Slosson Oral Reading Test*
- *Test of Adult Basic Education (TABE)*



Norm-referenced instruments are standardized so they are administered and scored consistently. Norm-referenced instruments allow raw scores to be converted into standard scores, as well as percentiles or normal curve equivalents, allowing for comparisons. (Other types of scores based on conversions of standard scores are grade-equivalents and developmental ages.) Norms provide a basis for interpreting performance in relation to defined population or groups.

In using a norm-referenced test, consider the following:

- Is the norm group appropriate and large enough for the individuals served by the project? For Even Start, comparison norm groups should reflect the age group(s) served by the project. The more defined the ages, the more useful the information about development will be, especially for young children. For example, the PLS-4 norms are based on approximately 15,000 children from birth through 6 years. The sample size of the norm groups must also be large enough to yield meaningful comparative data. An instrument may claim to use 2,100 children in the norming population, but the subgroup of children who are the same age as a project's sample may only contain 50 children. Information about the norming population and subgroups will be found in the test's technical manual.
- Are the correct scores being used and reported? Raw scores should not be reported. Most commercial instruments include scoring manuals that explain how to convert raw scores and interpret gain scores.

Criterion-referenced instruments. Criterion-referenced instruments are appropriate for measuring learning progress in specific content domains (e.g., reading, math). To measure outcomes for school-age children, schools typically use criterion-referenced tests in grades one through three. These instruments are best used to assess current knowledge at a point in time and to determine what the student needs to learn next. Criterion-referenced instruments measure mastery of skills or content against a specific set of performance standards. Criterion-referenced instruments can generate rich information about a student's progress and content mastery.

Criterion-referenced instruments compare student performance against pre-determined criteria, such as performance benchmarks, whereas norm-referenced instruments compare the student with other students.

In using criterion-referenced instruments, bear in mind that:

- The guidelines for scoring or understanding skill levels should be clear enough so that judging performance is consistent, fair, and accurate.

- In general, criterion-referenced instruments are more useful in measuring individual outcomes rather than group, or program, outcomes.

Validity and reliability. Validity and reliability refer to technical characteristics of assessment instruments; issues of validity and reliability arise with both norm-referenced and criterion-referenced instruments. An instrument has validity when it matches the construct it is intended to measure and predicts performance accurately. An instrument has reliability if it consistently measures the construct it was designed to measure.

Even if the state indicators specify the instrument, evaluators may want to know more about it. The technical manual will provide information about a test's validity and reliability. Below is a brief description of validity and reliability in relation to selecting an instrument to measure participant outcomes.

The **validity** of an instrument's score answers the question, "How do we know the instrument is really measuring what we want it to?" Measures are considered valid only if the instrument is used for the purpose for which it is designed. Validity factors include:

- Do project staff and relevant stakeholders consider the instrument valid? At a minimum, staff members must be able to look at the items on an assessment instrument and understand what is being measured. One criterion for validity is consistency with how the underlying concept is used in the field.
- Does the instrument permit generalization from one task to another? That is, can the instrument predict performance in a larger domain, such as reading or writing success?

The **reliability** of an instrument's score refers to the measure's consistency and repeatability. Scores are reliable to the extent they remain consistent despite fluctuations in the testing situation (i.e., time or day of week), or with different forms of the instrument, or different administrators.

Without reliability, it is difficult to know with any confidence what a student can do. One challenge in assessing young children is their rapid pace of development, making their behavior and performance inconsistent from day to day. Few instruments for assessing children have high reliability.

Another reliability issue concerns scoring consistency. Achieving consistency in scoring requires professional development and training scorers in applying criteria or procedures consistently. That is, to be reliable, one person scoring the instrument should get the same result as another scoring the same instrument.

Examples of criterion-referenced assessment instruments:

Children:

- *Get it. Got It. Go!*
- *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*
- *Phonological Awareness Literacy Screening (PALS)*

Adult:

- *Basic English Skills Test (BEST)*
- *Comprehensive Adult Student Assessment System (CASAS)*



Example: Using Data to Demonstrate Effectiveness

Indicator: 60% of adults who pre-tested at levels of 0-8.9 on the TABE and completed 100 hours of adult education instruction in reading will demonstrate one level gain as measured by the TABE post-test.

Number who completed 100 hours of adult education and took the TABE pre-test: 19

Number who demonstrated at least 1 grade level gain on the TABE post-test: 13

Calculation: $13/19 \times 100 = 68\%$

Finding: Program met expectation.



Role of project director. The project director collaborates with the evaluator to select assessment instruments. In some cases, staff members may be more familiar than the evaluator with instruments in certain content areas, such as adult education or early childhood education. In that case, the project director will suggest appropriate instruments.

The final selection of instruments is a mutual decision between the project director and the evaluator. Both bring expertise to the decision. Project directors may have content expertise and they know their staff's abilities. Evaluators will add expertise in assessing instruments and matching them with intended purposes.

Role of evaluator. The evaluator may help staff review the assortment of instruments used to measure outcomes. Are they adequate for the project's purposes? Are additional instruments needed for specific subgroups? Is an additional instrument needed to augment an instrument's validity? Can some instruments serve multiple purposes?

The evaluator's primary responsibility is to help interpret scores, but he or she also takes the lead in comparing various instruments and training (or overseeing the training of) staff members in how to use the instrument, how to score it, and how to convert scores. For example, if instrument scores are to be converted to percentiles, the evaluator needs to make sure staff members know how to convert raw scores and use percentile rankings. If the scoring conversion is especially complex, the evaluator may take responsibility for converting scores.

Using Outcome Data for Continuous Improvement

Projects typically demonstrate a program's success or effectiveness in relation to a standard or benchmark. For example, to claim success in helping adults improve their reading skills, a project's data must answer the question, "compared to what?"

The most common comparison for Even Start projects is between the performance indicator and actual results. (See chart, page 31, for an example of reporting outcome indicator data.)

Other comparisons to demonstrate a project's success to broader audiences include comparing a project's outcome data with a test's norms or with aggregated state results.

What if the outcome result is less than expected? Ideally, the findings will point to areas where change is possible, and the evaluator can conduct additional analyses. Two ways to do this are by making further comparisons and disaggregating group scores.

Making further comparisons. In addition to the comparisons listed above, other useful comparisons are looking at project scores over time and at pre- and post-test results for individual participants.

One benefit of required indicators is that project staff can compare similar outcome data over time. For example, a project may not have met the required state indicator, but its results show improvement over time. Staff can use this feedback to decide if current strategies are working.

Example: Comparing Indicator Data for Two Years	2001-2002	2002-2003
<i>Indicator</i>	<i>Program Results</i>	<i>Program Results</i>
<i>A1. 50% of the adults who have completed 100 hours of adult education instruction in reading and who pre-tested at Levels A-D on the READ or 0–8.9 on the TABE will demonstrate one grade level gain as measured by the READ or TABE post-tests.</i>	<i>59%</i>	<i>47%</i>
<i>C1. 40% of the adults participating in Even Start who have earning a high school diploma as their primary goal and who score 9.0 or above in reading and math on the TABE will earn a high school diploma or pass the GED during the program year.</i>	<i>21%</i>	<i>36%</i>
<i>Summary: Adults in group A1 achieved state indicator measures the first year, but did not in the second. Examining the reasons for this decline could be useful. Adults in group C1 did not meet state indicators either year, but the program showed progress from the first year to the second. It would be useful to learn what contributed to this increase in order to build on these factors.</i>		

Another approach compares pre- and post-test scores. This approach is typically used to determine if a change from one time to another is statistically significant—i.e., the results are unlikely to have occurred randomly.

Pre- and post-tests compare participants with themselves. The pre-test score shows what a student knew prior to enrolling in your Even Start project; the post-test score shows what he or she knows after exposure to the project or curriculum. The shift in learning from the pre-test to post-test may be a result of a program’s intervention, such as curriculum, instruction, etc.²

² There are many threats to validity associated with a pre-test and post-test design that may limit making statements of attribution. Because this Guide is not intended as a technical methods or statistics document, readers are encouraged to seek appropriate resources for a more thorough description of the limitations of pre-test and post-test analyses.

Taking sample size into account. Test results must be interpreted in light of the sample size. Typically, statistical tests are used with sample sizes of at least 30. For many Even Start programs, small sample size is a problem. Small sample sizes make it difficult to detect meaningful relationships and significant differences in learning. Findings from the statistical analysis may show no differences between groups when there really is a difference. One benefit of the pre-test and post-test design is that statistical t-tests are appropriate for small (fewer than 15) sample sizes. Another solution is to collect longitudinal data and study patterns over time. Another is to collapse groups in order to find patterns. For example, the results for several age groups such as 2 to 5 year olds can be combined.

Disaggregation by group scores. Another way to analyze outcome data is to sort data within the project population by meaningful sub-groups, such as participants who achieved the outcome and those who did not. These sub-groups may then be compared on relevant factors such as demographic characteristics (e.g., age, gender, race/ethnicity), contact hours, duration, entry level scores, opportunities for learning, etc.

Interpreting data. When interpreting data, the purpose of the analysis should be central. The evaluator and program staff will determine if the data are meaningful. The local evaluator arranges data so project staff can understand the findings and interpret the results. Chapter 11 presents approaches for reporting data. Below are issues evaluators should consider when interpreting data.

- Exercise caution in attributing results to causes. Although findings may show achievement gains, it may be hard to attribute them to the Even Start project. Other factors may have contributed to participants' achievements. Program staff and evaluators need to interpret findings with care and limit their conclusions to those the data can support. Using several data sources and methods can help rule out alternative explanations and demonstrate the project's influence.
- Statistical significance is not necessarily practical significance; statistically significant findings may have no real practical value. For example, findings may be statistically significant when there has only been a very small change in scores, such as $\frac{1}{2}$ point. There may be a gain on a few test items, but the gains may not reflect meaningful progress.
- When making comparisons over time, the evaluator should compare the same participants, with the same intervention, over the same time period.

Armed with outcome data he or she feels confident about, the evaluator will likely next meet with staff members, present the findings, and guide discussions about what the findings mean for the project. Ideally, interpretations of outcome data should point to program areas that can be improved, such as professional development, curriculum integration, or the intensity of information. Some changes may be significant enough to become the focus of a new evaluation. Others may be less important, but all proposed changes should be documented so that their effects on participant outcomes can be tracked.

With the evaluator, the program leader should facilitate a staff discussion about understanding and using the data, and the project director should use the data consistently to make decisions, particularly about resources and program changes. Using data frequently, the project director can help staff members see the benefit of collecting data—an important step since staff will usually collect it.



Chapter 4

Refining

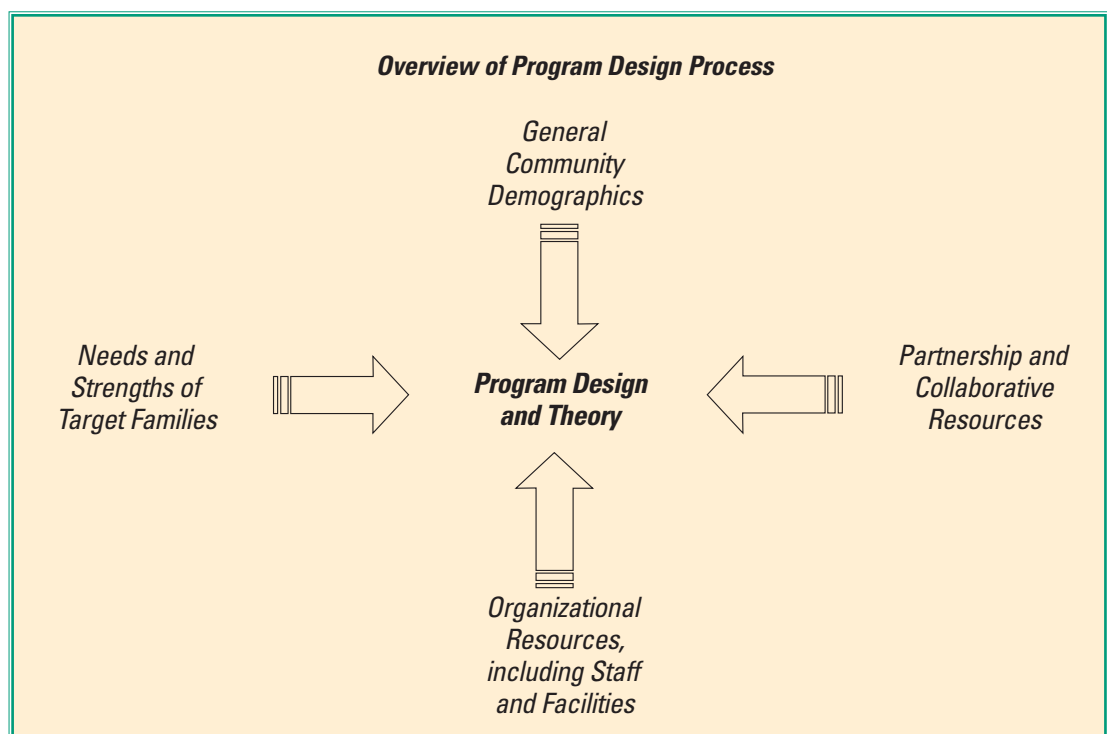
the Program Design

Evaluation can play an important role in program design, especially for first-year projects and those that have been identified as needing improvement. Chapter 4 addresses how evaluation activities can inform program design. It covers:

- target populations;
- partnerships and collaborations;
- staffing and facilities; and
- program theory.

The first step in project design is to understand the project's context. Here context means the target community's demographics, the needs and strengths of target families, the nature of local partnerships and collaborations, and Even Start organizational resources such as staffing and facilities. The program evaluator helps project staff determine families' service delivery needs, the resources available to meet them, and the effectiveness of program services. These findings organize the program design.

Program theory sets the parameters of program design. Program theory consists of the assumptions and presuppositions that underlie program practices and distinguish family literacy from other educational approaches. These assumptions and presuppositions are a set of logical statements that serve as the theory, or rationale, for program design. Evaluators and program staff need to ensure a coherent relationship between the program design and intended outcomes. The figure below depicts the program design process.



Faced with actually making the approved program design “come to life,” project staff may find that they need to refine their original designs or even modify them to better match the context in which the program will take place. Making major modifications in design should be done with the assistance of information from the local evaluator—and with the clear understanding that local projects must receive permission from the state agency before changing approved program designs that were part of the initial proposal.

Target Populations

Project staff members need to maintain a thorough knowledge of the area’s demographics, including poverty levels, languages spoken, and population density. This information

can usually be assembled from local service providers, census information, or geographical information systems (GIS), which are often made available on the World Wide Web. Project staff should re-examine their target area’s demographics from time to time because populations can change.

The federal statute governing Even Start requires programs to serve:

families most in need of services . . . as indicated by a low level of income, a low level of adult literacy or English language proficiency of the eligible parent or parents, and other need-related indicators (Section 1235(1), ESEA).

Some Even Start programs adopt a family literacy delivery model without a sufficient understanding of the targeted population, with the result that the program design does not match the needs of the target population. Project staff members need to ask specific questions throughout the life of a program—during both planning and implementation—to ensure that the program design actually addresses the target population’s needs.

Participation criteria. Once project staff have defined the target population, the next step is to develop criteria by which families who meet that definition will be recruited. This requires the input of key stakeholders, such as community representatives, current service providers, and educators. Family selection criteria are usually developed in concert with a review of organizational and community resources (see sections on partnerships and collaboration, page 40). Project staff members need to re-visit this process periodically to make sure they are meeting the intent of the statute by recruiting and serving families most in need of services.

By statute, the project must consider income and literacy status in identifying those families most in need. Some projects develop additional criteria to further define the subset of families they are going to serve (teen parents, English language learners, etc). Criteria must be clearly defined.

Example: Understanding Participants’ Constraints

In its grant proposal, an Even Start project planned to implement a program requiring adults to participate in morning and afternoon adult education and parenting classes. Though there was initially a strong demand for adult education classes, actual attendance was poor during certain months of the first year. On further investigation, the project evaluator realized that most of the families were migrant farm workers; their work schedules prevented them from attending daytime classes during harvest times. The program staff made the necessary schedule changes, and adult participation in the program increased.



Participant selection criteria could take the form of an instrument that reflects a program’s service priorities and allows families to be ranked in order of need. Staff members should agree on the scoring and interpretation of a participant selection instrument. The example below shows a measurable, criterion-based family selection instrument that some projects use.

Sample Family Enrollment and Prioritization Form						
<i>Indicators</i>	<i>Level Of Need Related To</i>					
	<i>0 N/A</i>	<i>1 Low</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5 High</i>
<i>1. Poverty *</i>						
<i>2. Literacy Level of Adult*</i>						
<i>3. English Proficiency*</i>						
<i>4. Employment Status</i>						
<i>5. Homelessness</i>						
<i>6. Handicapping Condition</i>						

**These criteria must be used to determine those most in need. Other criteria are optional.*

Once program staff members have a clear understanding of the target population and the type of families to be recruited, they next assess those families’ specific needs. For new projects, the assessment will target the population of potential families. The staff of ongoing projects should periodically re-assess families who are or were in the program.

Assessments can include surveys, one-to-one interviews, and focus groups; each has its strengths and weaknesses. The evaluator should ensure that the project uses appropriate assessments to determine the target population’s needs, including screening assessments to determine language proficiency and skill level. Most projects administer formal assessments to adults once enrolled in order to determine the appropriate level of instruction.

Recommendations on Using Surveys with Adults with Limited Literacy Skills

- Questions are easy to understand and in the language(s) of the target population.
- The number of questions is as small as possible.
- The survey uses large print and is easy to read.
- Survey response items are unambiguous and easy to mark.
- Dichotomous (yes/no) or easy-to-use Likert scales are used.

Using surveys and focus groups. Program staff can assess families’ literacy needs through a survey or questionnaire. A survey targeting adults with limited literacy skills should be easy to read and understand; survey questions should address the types of services needed by families.

Survey questions can garner information on when and where services should be offered. The example on page 39 illustrates a survey some Even Start projects use to ensure their activities are offered when families can participate.

Sample Family Literacy Scheduling Survey

Directions: Please place a ✓ in all the boxes that of times and events you and your children could likely attend. (You may choose more than one time.)

Activities	<i>Morning 8 – 12 AM</i>	<i>Afternoon 1 – 3 PM</i>	<i>After school 3 – 5 PM</i>	<i>Evening 6 – 9 PM</i>
<i>Adult Education Programs</i>				
<i>Parent Education Activities</i>				
<i>Infant Time</i>				
<i>Preschool</i>				
<i>School-age Activities</i>				
<i>Interactive Literacy Activities between parents and children</i>				

Some adults may not have sufficient literacy skills to complete a survey. For this reason, projects may decide to use one-to-one or focus group interviews. One-to-one interviews have the advantage of targeting key informants (people with extensive information about the community) or potential participants with open-ended questions. The obvious disadvantages of one-to-one interviews are the considerable time individual interviews take and the limits on generalizing the results to the whole target population.

Both during project start-up and at regular intervals, project staff should use one-to-one interviews with key informants, including local literacy providers. They can offer a valuable perspective based on their own experience of providing services to the target population.

Focus groups usually contain no more than eight participants who are asked a series of open-ended questions. Their responses are recorded and later analyzed. Focus groups give the interviewer the luxury of asking follow-up questions to clarify responses. However, focus groups face many of the same limitations one-to-one interviews do, and focus groups may give skewed results if an inexperienced interviewer allows one or two interviewees to dominate the conversation. Inexperienced program staff members should seek the evaluator's expertise when using one-to-one or focus group interviews.

The following sample questions could be used in one-to-one and focus group interviews:

- How did you hear about Even Start?
- What, if any, changes should be made in the early childhood, adult literacy, and parent education to better meet your needs?
- Are there any barriers that prevent you and your family from receiving services (prompts: language, access, scheduling, support services)?
- What changes would make the program better?

“Fatal flaws” common to low-performing projects.

- *Model: Program models have unmanageable sites in terms of quantity or locations, small numbers of participants, or are missing program elements.*
- *Partnership/leadership: Programs demonstrate no functional partnership, lack clear roles and responsibilities for staff, or provide no support for the coordinator.*
- *Director: Directors do not have enough time allocated to the project, lack skills and experience necessary for the managerial role, or lack control of the Even Start budget.*



Partnerships and Collaborations

Both partnerships and collaborations support Even Start programs. Partnerships are formal relationships among one or more LEAs and one or more other non-profit public or private entities. Partnerships are eligible to apply for Even Start sub-grants. Collaborators may provide services, but are not formal partners in the “eligible entity” sense. They may be informal and have little bearing on day-to-day operations, or elaborate and require extensive commitments of time and resources. Both partnerships and collaborations are essential to avoid service duplications and to ensure long-term Even Start program sustainability.

Project staff should ask potential partners and collaborators for relevant service and evaluation reports and thoroughly examine the services proposed partners and collaborators have provided before making formal agreements. This is an important step to ensure that services offered to Even Start families are of high quality. This examination will likely require experts, including evaluators, to judge both the quality of services offered and the extent to which they meet the target population’s needs. Even Start evaluators are independent of the program and are therefore uniquely positioned as “outsiders” to make recommendations that affect program delivery. Their independence is vital to maintaining objectivity in determining service quality.

All agreements with partners and collaborators should include a commitment to provide evaluation data and specifics about how services will be evaluated. These agreements should also reflect a mutual understanding of the curricula and assessments to be used as well as standards to judge program success. Ongoing programs need to ask each of their partners and collaborators the critical questions, *Are services mutually planned and coordinated? Do organizational partners and collaborators provide high quality services to Even Start clients? How can services be improved?*

Many projects formally evaluate their partnerships and collaborations each year. This is helpful in determining both service quality and the extent to which projects are building sustainable family literacy programs. The example on page 41 shows an instrument that could be used to determine the strength of partnerships and collaborations.

Family Literacy Interagency Service and Resource Measure			
<i>Name of Partner or Collaborating Agency:</i>	<i>Definitive Yes</i>	<i>Partially (to a limited extent)</i>	<i>Definitive No</i>
	3	2	1
A. Getting Started			
1. <i>Organization shares a detailed understanding of the Even Start project's goals and objectives.</i>			
2. <i>The organization shares the Even Start vision.</i>			
B. Service Provision			
1. <i>There is a formal agreement with the Even Start fiscal agent.</i>			
2. <i>There are specific, agreed-upon indicators by which the quality of services will be determined.</i>			
C. Evaluation			
1. <i>The collaborating agency agrees to share evaluation data.</i>			
2. <i>Meetings will be regularly scheduled to evaluate the success of the partnership or collaboration.</i>			
3. <i>A continuous improvement process addresses program weaknesses, and the organization is committed to participate in Even Start sponsored professional development activities.</i>			

The federal statute requires Even Start projects to be operated by a partnership of one or more local educational agencies (LEA) and non-profit entities. Thus, Even Start projects usually operate within larger organizations' administrative structures and depend on them to varying degrees. The resources and administrative structures these organizations provide necessarily influence program implementation. A thorough examination of such an organization's resources, including administrative structure, staffing, and facilities is an important part of every evaluation. Important evaluation questions are:

- *How does the organization support the Even Start program?*
- *What internal barriers might impede the Even Start program?*
- *What can be done to make the Even Start program more successful?*

Staffing and Facilities

Hiring and training staff are key administrative functions of every Even Start project. Staffing is a principal variable affecting project quality, and local projects should dedicate resources to determine the qualifications and expertise of their staff in evaluating professional development activities.

Every state is mandated to implement minimum staffing qualifications for Even Start projects. Each local project must determine if its staff needs additional competencies to serve its target population effectively. For example, programs serving second-language populations may need instructional staff with certain linguistic capabilities. Because personnel quality is so important, projects should keep detailed records of staff qualifications. The example below is a personnel instrument for monitoring staff quality.

Quality of Personnel Record					
<i>Position/ Name</i>	<i>Degrees Certificates and Credentials</i>	<i>Years of Experience Working in Family Literacy</i>	<i>Specialized Family Literacy Training</i>	<i>Language Proficiency</i>	<i>Time Devoted to Working With Even Start Families</i>

It has been said that facilities drive programs—that is, facilities affect program delivery and design. This is especially true for family literacy programs, which address the instructional needs of adults, school-age children, preschoolers, and infant/toddlers. Some programs have adequate space; others struggle to secure sufficient space. The amount and quality of space are important, and often overlooked, variables which can deeply affect program quality. The key terms here are appropriateness and accessibility. Key questions include:

- *Does the project have sufficient space to carry out its program activities?*
- *Is the space appropriate to the ages of participants?*
- *Do the instructional environments encourage literacy development?*
- *Do the facilities encourage program integration and intergenerational activities?*
- *Are the facilities accessible to parents and children with disabilities?*

The location of Even Start facilities is equally important. Even Start programs are offered in various locations, including schools, community centers, homes, churches, and prisons. Programs need to ask: *Are the facilities close enough to the target families? Is the staff located at the same site where services are offered? Is the site within walking distance of families with young children? If not, is free or low cost transportation available, or can the project provide it?*

Many Even Start programs offer services at more than one site. Some offer adult classes at one site and serve children at another site that may be a great distance away. Further, some rural Even Start projects offer services miles away from where families live. For them, an evaluation of program accessibility is especially in order. With the help of an evaluator, project staff may need to explore the feasibility of designing programs that feature alternative service delivery such as distance learning or additional home-based services.

Program Theory

Program theory is the set of presuppositions, or logical statements, that serve as the rationale for the program and distinguish it from other educational approaches. A program's design and practices grow out of these underlying presuppositions.

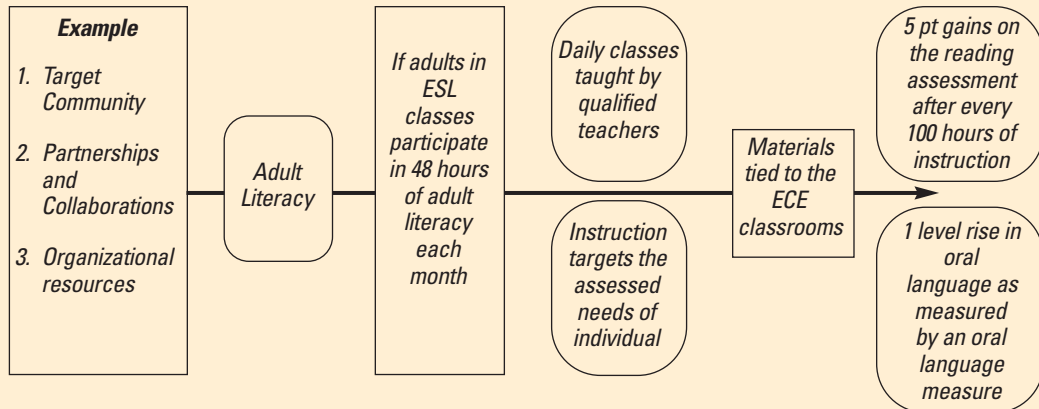
The explicit theory of family literacy is that parents and their children engaged in concurrent, coordinated learning will achieve greater literacy outcomes. Family literacy involves more than the components of adult education, parent education, early childhood education, and interactive literacy activities between parents and children. Family literacy presupposes that the parallel involvement of parents and children in joint learning promotes greater literacy for both parents and children than if they were not mutually involved in the learning experience.

This theory of family literacy plays out in different ways at the local level, specifically in how programs integrate adult education, parent education, early childhood education, and interactive literacy activities between parents and children. Project staff members need to choose or develop a model of integration (plans for how the four components interact) that is best adapted to meet contextual realities.

The evaluator can also help projects recognize other, implicit, theories or beliefs inherent in the program's design, such as "Reinforcement of instruction in the home is necessary for all four core instructional components." The evaluator can help project staff make these beliefs explicit. To this end, many programs develop a logic model—a way of visualizing how the program design features interact. Developing a logic model can help connect context, design, resources, and outcomes and uncover the implicit beliefs that underlie the logic of the program design. There are several different types of logic models. The figure on page 44 presents a template for a simple logic model some family literacy projects use.

Template for a Family Literacy Logic Model

Context + Components + Assumptions + Component Activities + Inter-Component Integration Activities = Outcomes



Note: This example is completed for only one component; a project logic model would also include the parenting education, interactive literacy activity, and early childhood education components.

Project directors are encouraged to engage their partners and collaborators in assessing partnerships and collaborations to ensure the best use of resources. Staff members should understand how current administrative structures, staffing, and facilities affect their program so they can suggest changes as needed.

Staff members should understand how their project’s design will ultimately lead to services that produce literacy gains for children and adults. There should be an explicit link between program context, implementation, instructional activities, and anticipated outcomes. The project may need to revisit the design if program outcomes are not achieved.



Chapter 5

**Program Content and
Implementation**

Implementation is program design in action. Evaluating program content and implementation means determining that a program is executed as originally designed and meets a high standard of quality. Often called process or formative evaluation, this continuous evaluation assumes an explicit program design, a mutually agreed upon set of quality criteria, and faithful data collection. Chapter 5 examines evaluation as it applies to:

- curriculum and instruction;
- professional development;
- service delivery; and
- quality criteria for program self-assessment.

Curriculum and Instruction

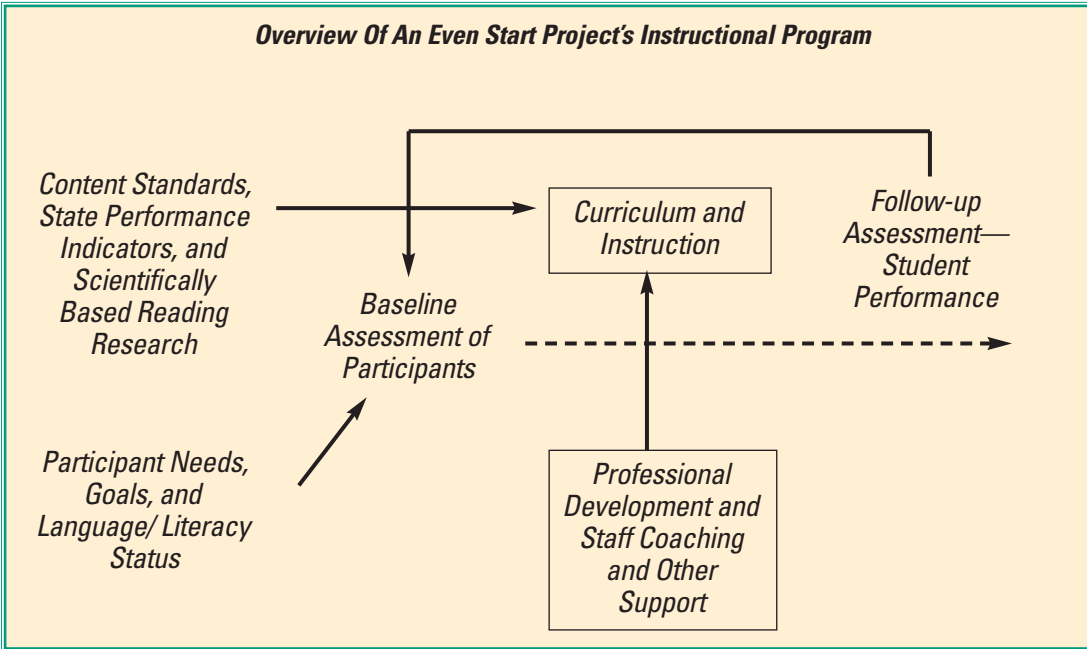
Curriculum and instruction are at the core of program implementation. The legislation states that the purpose of Even Start is:

to promote the academic achievement of children and adults; assist children and adults from low-income families to achieve to challenging State content standards and challenging State student performance standards; and use instructional programs based on scientifically based reading research and the prevention of reading difficulties for children and adults, to the extent such research is available. (Section 1201, ESEA [20 U.S.C. 6361] of the ESEA.)

To evaluate curriculum and instruction properly, it is helpful to grasp the complex interrelations among participants' needs and goals, content expectations, curriculum (the objectives, content materials, and sequence of materials), instruction (pedagogy and methods), assessments of learning, and staff preparation (professional development and support). The figure on page 47 presents a simplified overview of the process.

The Even Start legislation mandates that programs use instruction based on scientifically based reading research to help participants achieve state content and achievement standards. These standards serve as a roadmap for the instructional program. Four important questions are:

- *How does the program's curriculum address relevant content and achievement standards?*
- *Does the curriculum for the core instructional components use scientifically based reading research?*
- *How comprehensive is the curriculum? Is it properly sequenced?*
- *Is implementation faithful to the intended curriculum delivery?*



Participants' needs, goals, and language/literacy status. A proper starting place in evaluating curriculum and instruction is the participants themselves. The principal question is: *How does the program match students' development and learning needs?* The answer will require a clear assessment of participants' needs, goals, and educational status. Even Start programs should offer educational interventions at the instructional level of their participants.

The evaluator helps determine how well the instructional program meets participants' needs. Evaluators often use focus groups or surveys to ask participants directly if the program is offering them the services they need. Other options include direct observation of instruction and expert review of curricula and materials. Evaluators should share their findings directly with the instructional staff; together they should develop a plan to improve instruction. Evaluators can follow this up by assessing how well the improvement plan is being implemented.

Appropriateness and fidelity of implementation. Evaluators are especially concerned with the appropriateness and fidelity of instructional delivery. Appropriateness of delivery is the extent to which the instructional program matches the development and learning needs of students. Important questions are:

- *How does the instructional program engage the interest of students?*
- *Does the program provide the instructional support necessary for students to meet their goals and targeted standards?*
- *Are the cultural and English language acquisition needs of students addressed within the instructional framework?*

Fidelity of delivery is the extent to which the proposed curricula are actually used as originally envisioned. Invariably, teachers change curricula to fit their instructional programs. Do these changes add or subtract valuable content that

students will need to know? The project evaluator may have to work back from assessment results to the instructional program to learn if changes in curricula are affecting student performance. Other important evaluation questions include:

- *Is the instructional pedagogy appropriate?*
- *How does the instructional environment support learning?*
- *How well-integrated are instructional approaches for the different instructional components?*

Assessment. Assessments used to place participants must align with the instructional program and measure student achievement of specific learning milestones or constructs. Assessments may include student portfolios or observational records in addition to standardized instruments. Chapter 3 of this Guide discusses assessment in more depth. Evaluation can focus on the relationship between instructional programs and the project's assessments. Key questions are:

- *How well are assessments aligned with the project's instructional program?*
- *How are assessments used to inform instruction?*
- *Do the assessments evaluate local/state standards at appropriate levels of difficulty?*

Professional Development

Professional development is basic to any Even Start project, and all Even Start projects must provide staff training. As defined under the No Child Left Behind Act, professional development means high quality, sustained, intensive, and classroom-focused activities that are designed to have a positive and lasting impact on teachers' knowledge of their academic subjects, their classroom instruction, and their understanding and use of scientifically based instructional strategies (Section 9101(34) [20 U.S.C. 7801(34)]).

Evaluating professional development takes place both at the program planning and implementation phases. Initially, evaluations are concerned with assessing staff needs, while the second phase of evaluation concerns the appropriateness and utility of professional development activities themselves.

Project staff members are a great source of information in planning professional development programs. They are in the best position to know what additional training they need in order to better meet the demands of their job. Many projects regularly and formally assess staff needs. The example on page 49 is excerpted from a professional development questionnaire some Even Start projects use annually to plan professional development activities:

Professional Development Self-assessment

1. What is your future professional goal?

2. What additional professional development do you need to be successful in your current position?

3. Family literacy programs are complex; they require skills in several different areas. For each item below, circle the number to the right that best describes your knowledge or ability.

Items	Poor	Fair	Good	Very Good	Excellent
1. Knowledge of family literacy	1	2	3	4	5
2. Ability to recruit and enroll families	1	2	3	4	5
3. Knowledge of scientifically based research related to the project's instructional programs	1	2	3	4	5
4. Using data to improve program design and instruction	1	2	3	4	5
5. Knowledge of program evaluation and state quality indicators	1	2	3	4	5

Many Even Start projects regularly assess their professional development activities through evaluation forms distributed at the end of each professional development activity. Staff members review this information to determine the value of the activity and whether professional development activities need to be modified.

Professional development can also be the subject of a focused inquiry. Such an inquiry could examine how strategies presented in professional development activities are incorporated in the instructional setting. For example, a focused inquiry could explore how preschool teachers who attended training in scientifically based reading research integrated what they learned into their daily teaching.

Focused inquiries into professional development activities may use observational assessments that target the use of specific desired behaviors in instructional contexts. These observational assessments are most effective when they are used repeatedly by trained observers in the same classroom so that they reliably capture the instructional program. Other data sources that may demonstrate instructional integration include portfolios, student work, and lesson plans.

Example: Examining Service Delivery

One Even Start project offered an after-school program at two school sites for all of its school aged children. An analysis of assessment data showed that students at one site performed significantly better than students at the other site. Further examination of the after-school programs found that the program serving the higher performing students mandated participation in daily homework and reading clubs, while the under-performing after-school program was largely recreational. The Even Start project added daily homework and reading clubs to the site with low performing students.



Service Delivery

A global term, service delivery encompasses all the activities in the program design that support implementation of the family literacy program. These include developing classes and coordinating support services, scheduling, staffing, and administration.

The evaluation of service delivery is concerned with fidelity—the extent to which the program design has been faithfully implemented. The first step is to define the program services. These definitions should be specific and correspond to the activities contained in the project’s logic model. Staff members, collaborators, and participating families should all know these definitions. For example, all stakeholders in a program serving second language learners should know what qualifies an adult for English as second language (ESL) classes, what ESL classes offer, and what criteria determine when a student moves out of ESL into regular adult basic education classes.

A shared understanding of the particulars of program services is especially important in Even Start projects. By their very nature, interactive literacy activities between parents and children combine both parenting and early childhood education. Activities may be designed primarily for parents, children, or both, making it essential that staff members and parents know the focus of services. Data collection and reporting errors are likely if staff members do not share an understanding of how services are defined and recorded.

Effective evaluation depends on a clear understanding of the program design and delivery system. Variability in service delivery usually complicates evaluations in highly individualized programs or programs that operate in many sites. It is difficult to make general statements of program quality when the program’s defining characteristic is variability.

Intensity is another important concept in evaluating service delivery. Intensity concerns the extent to which project participants receive the program as designed. Both service location and scheduling can influence intensity. Successful programs provide enough services (intensity) so that participants can meet their goals. The issue of equality is central to intensity. See Chapter 6 for more on program intensity. Two important questions are:

- *How much instruction does a family need in order to meet the program outcomes?*
- *At what point does staff conclude that an individual or family has not received an Even Start intervention because they did not participate in enough services?*

Quality Criteria for Program Self-Assessment

Many states have developed quality criteria for adult education, early childhood education, and parenting programs. Local programs can use these criteria to guide their implementation or develop their own quality criteria to address areas of specific concern such as the quality of content and implementation.

Locally developed criteria should be grounded in existing research and best practices. Criteria should be clearly and unambiguously written and concern measurable aspects of program quality. In creating high-quality criteria, it is important to remember that family literacy is greater than the sum of its major components—more than just adult education, early childhood education, interactive literacy activities, and parenting education. Criteria should include component integration, recruitment and retention, administration, and collaboration.

Some projects have formalized the assessment of how well their programs meet quality criteria by creating their own instruments or using an established criterion-based instrument. These instruments often include a scale by which stakeholders rate how well each criterion describes the project. Evaluators can use these results to identify areas where greater focus is needed for program improvement. The example below is a sample set of criteria for one aspect of program operations.

Sample Set of Quality Criteria for Program Administration		
Component: Project Administration Indicators	Average Score (1=Not Descriptive 5= Very Descriptive)	Comments
<i>1. The leadership has an articulated vision of family literacy for the program and the role of the community in supporting this vision.</i>		<u>Strengths:</u> <u>Enhancements:</u>
<i>2. Program administrators' management styles foster shared decision-making and team-building that involve partners, collaborators, staff, and parents.</i>		<u>Strengths:</u> <u>Enhancements:</u>
<i>3. Program administrators use a management and accountability system that allows them to measure program effectiveness and outcomes.</i>		<u>Strengths:</u> <u>Enhancements:</u>
<i>4. The program has a written staffing plan that includes job descriptions and expectations for each position.</i>		<u>Strengths:</u> <u>Enhancements:</u>

The point of collecting information about program quality is to improve the program. Stakeholders are encouraged to compare their quality ratings, discuss their differences, and create a plan of action that leads to program improvement. Program leaders should repeat this process at least annually, creating new plans of action as necessary. The example below shows a sample plan of action.

<i>Sample Plan of Action</i>					
<i>Desired Outcomes</i>	<i>Actions to Take</i>	<i>Resources Needed</i>	<i>Responsibility</i>	<i>Timelines</i>	<i>Progress to Date</i>

Together with their evaluators, project staff should review their progress in following the plan of action. The plans themselves may call for changes in program design or implementation. Such changes can be time-consuming and may necessitate “re-thinking” a program’s basic elements. It may take courage to replace collaborators, change services, or redirect resources – but this may be the appropriate outcome of a meaningful evaluation. Project directors will want to check with their state agency on any proposed design changes.

The evaluator’s role is to guide the project in making data-driven decisions to deliver the best services possible. Such decisions require the systematic collection of data on the key areas of implementation: curriculum/instruction, professional development, and service delivery. Together, evaluators and project staff use data for planning, assessing quality, and effecting continuous improvement. Evidence-based evaluation will lead to meaningful, positive changes for the project and for families.



Chapter 6

**Program Intensity and
Participant Attendance**

The local evaluator monitors the systematic collection of participation data and provides analyses useful in both interpreting participant outcomes and guiding the continuous improvement of program services. This chapter covers:

- key evaluation questions related to participation;
- participation variables;
- managing and analyzing data; and
- using participation data.

Participation data document the number of hours of services offered in the four core instructional components (program intensity) and the number of hours that participants attended those services (participant attendance). These data help determine whether projects are offering enough instructional hours and services to ensure that participants can achieve their goals, and whether families are participating at a level such that they can achieve their goals.

Instructional opportunities and participation affect participant outcomes directly. Participation data are particularly relevant given the nature of Even Start programs and the populations served—i.e., the “most in need” families, who frequently face problems that compromise their ability to participate in program services regularly (e.g., work pressures, substance abuse, lack of housing, etc.). Keeping accurate attendance records and monitoring family participation patterns can be difficult because Even Start families tend to enter and exit programs throughout the project year. Ensuring that projects offer enough instructional hours in the four core components has historically proved challenging for project staff. For these and other reasons, participation data are considered critical for local evaluations and constitute an essential element in the *Framework for Planning Local Evaluations*.

Key Evaluation Questions Related to Participation

The local evaluation documents the extent to which projects offer services at the level of intensity participants need to achieve their program goals. Key questions are:

- *Do projects offer enough hours of service for participants to achieve outcomes?*
- *Are program services of sufficient duration for participants to achieve outcomes?*

Answering questions about program intensity is just one aspect of participation data. The next set of evaluation questions looks at attendance patterns, or participation in the services offered:

- *Are participants attending enough hours of service to realize outcomes?*
- *Are participants staying in programs long enough to achieve their goals?*

“Participant attendance” reflects the actual number of hours that individuals participated in program services and the length of time the family actively participated in the program.

Analyzing participation data can also support efforts to improve program services. Key evaluation questions that can inform continuous program improvement include:

- *How are families meeting the project’s participation expectations and attendance policies across the four core instructional components?*
- *Do participation patterns differ by program component?*
- *Is the project recruiting families who are committed to program participation and likely to remain in the program long enough to achieve their goals?*

Participation data provide evidence of participants’ 1) satisfaction with program services, 2) understanding and compliance with project participation policies, and 3) commitment to participating in a family literacy program. Data documenting participation patterns and attendance rates that fall below project expectations may spur staff to gather additional data on program quality, the accessibility and scheduling of services, the recruitment of appropriate families, staffing concerns, and so forth.

Linking program intensity and participant outcomes. The premise of Even Start family literacy is that families “most in need” will participate in a family literacy program at a level of intensity that will enable adults to improve their literacy skills, become partners in their children’s education, and ensure that children will be prepared for success in school and life experiences. Findings from the second national evaluation study of Even Start family literacy

programs identified family participation as an important variable in achieving participant outcomes. The conclusion that “the intensity of program services and the duration of program participation are correlated with participant achievement” led to an amendment of federal Even Start legislation and the definition of the term “family literacy services.” This definition directs

Even Start programs to provide services at the level of intensity needed to ensure that families can meet their program goals.

**Partial Definition of Family Literacy Services
(Section 9101 (20) of the ESEA)**

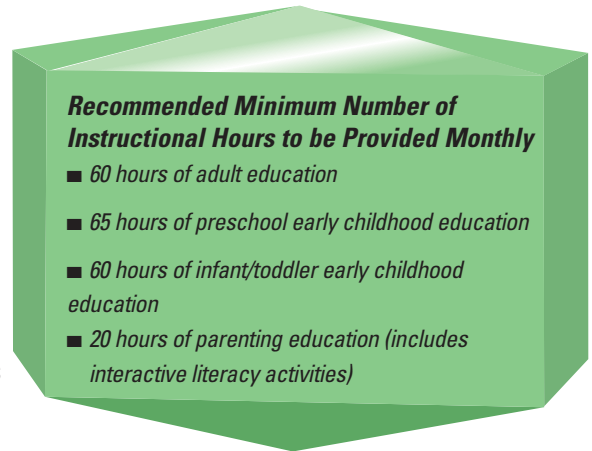
“...services provided to participants on a voluntary basis that are of sufficient intensity in terms of hours, and of sufficient duration, to make sustainable changes in a family...”

Example: Problems with Participant Attendance

Situation: *The Even Start program offered GED preparation classes three hours a day, five days a week. Although the classes started at 8:00 a.m., most parents arrived between 8:30 a.m. and 9:00 a.m.; several parents left early to pick up their children from kindergarten. Adult attendance was sporadic and the Even Start staff considered it a good day when half of the participants showed up. At the end of the year, many adults expressed dissatisfaction with the program because they did not pass the GED examination.*

Conclusion: *Although the program offered 60 hours of adult education services a month, participants’ attendance records showed that adults participated in fewer than half the instructional services offered. Adult attendance was irregular, with long periods of absence interspersed with weeks of steady attendance. Staff concluded that they needed to communicate better and enforce the 80% attendance policy, develop incentives to encourage regular attendance, and consider changing program hours to better accommodate parents’ schedules.*

Increased program intensity leads to increased participant attendance. The second and third national evaluations of Even Start programs examined the relationships among program intensity, attendance, and achievement of participant outcomes. Evidence that families who participated in high-quality intensive instructional programs were more likely to achieve their goals led to recommendations for minimum intensity levels in the required core instructional components of adult literacy/education, parenting education, early childhood education, and parent/child interactive literacy activities. The recommendations are based on the number of instructional hours offered by the Even Start projects which had the highest program intensity levels in the third national Even Start evaluation. The Department recommends a minimum number of instructional hours for Even Start programs (see right).



Linking family attendance rates and participant outcomes. Findings from the third national evaluation study revealed that “families in the program do not take full advantage of the services offered by Even Start projects; they participate in a small amount of instruction relative to their needs and program goals.” Analyses of Even Start family participation patterns (2000-2001) showed that, on average, families received instructional services for approximately seven months and attended fewer than half of the instructional hours offered in the four components. These findings suggest that Even Start programs need to be more active in developing strategies to increase participant attendance and encourage families to remain in programs for longer periods of time. In 2000, the Even Start statute added Attendance and Retention as a required program element.

Evaluations of Even Start programs can support program efforts to address the required program element of Attendance and Retention by 1) designing and monitoring systems to collect participation data, 2) determining the effectiveness of strategies to enhance participation, and 3) providing projects with ongoing analyses of family participation patterns.

Participation Variables

Four participation variables can be calculated to construct the most complete picture of participation data. They are:

Program Intensity. The number of instructional hours offered in each component during the project year;

Participant Attendance. The number of hours a participant attended the services offered in each component;

Attendance Rate. A percentage based on the number of hours the participant attended out of the number of hours possible for the participant to attend.

Duration or Length of Enrollment. The length of time that a participant remained an active participant in the program.

Determining activities to be included in participation calculations. Even Start projects offer a variety of services and classes within the four core instructional components. An important first

step is to identify those instructional activities that should be counted in participation calculations.

For example, parent participation in a substance abuse counseling group should be counted as support service hours offered by the program, but would

not be included in the participation calculations for parent education

because substance abuse counseling is not considered an instructional service. Parent education activities that should be counted in the participation calculation are those that have an instructional/literacy focus (e.g., child development, literacy development). Project directors and evaluators should review state guidelines (if provided) and work together to determine the specific instructional activities that should be included in participation calculations for Even Start.

Program intensity. Program intensity represents the total number of instructional hours offered during the project year. Program intensity hours are calculated for the four core components, including all program services identified as appropriate for instruction (e.g., center-based classes, individual tutorials, home-based programs).

Program intensity data demonstrate whether projects are offering services of sufficient intensity to help families achieve their goals. Instructional activities may be provided by Even Start or by collaborating agencies. Because the number of hours offered each week can vary and programs close for school holidays, staff development days, etc., project staff should document the actual number of instructional hours offered rather than estimates of hours based on weekly schedules (e.g., GED classes are offered three times a week). To determine program intensity accurately, project staff must carefully maintain records and document the number of instructional hours offered weekly.

Calculating program intensity. Calculating program intensity simply consists of adding the total number of instructional hours offered over the course of the project year. Program intensity is the

Example: Activities Counted as Parent/Child Interactive Literacy Activities

Interactive Literacy Activities have an instructional/literacy focus, involve both the parent and the child, and are supervised or structured by Even Start staff or teachers, such as:

- *Parents reading books with their children and*
- *Field trips designed for parents and children to learn together.*

Activities that should not be counted as Interactive Literacy Activities instruction include:

- *activities that do not engage both parent and child,*
- *home-based activities that are not supervised or structured by program staff (e.g., parent reads to child at home but does not engage in debriefing or follow-up activity on reading experience, parent watches TV with child but does not participate in a structured follow-up activity), or*
- *free-play outdoor activities that do not have an educational focus.*

(Excerpt from California recommendations to local projects)

Attendance and Retention

Each Even Start project must “encourage participating families to attend regularly and to remain in the program a sufficient time to meet their program goals” (Section 1235(11) of the ESEA). Each program determines what is reasonable attendance for a family in order to achieve its educational goals and to remain active participants in the program. If a family is not attending regularly, the project needs to work with family members to determine what they need in order to participate actively. If such attempts still fail, the project should transition the family out of Even Start, perhaps to a more suitable program, and recruit a more committed family.

number of hours of instruction that the average participant could receive or the number of hours that programs expect participants to receive if they attended 100% of the time and were enrolled for the entire project year.

Example: Calculating Program Intensity³

Early Childhood Education: Preschool-age children attend preschool for four hours a day for 180 days during the school year and also receive 60 hours of ECE during an Even Start summer enrichment program. Children enrolled in K-2 may participate in after-school supplemental tutoring programs (2 hrs/day for 80 days), attend Even Start during their off-track sessions (3 hrs/day for 32 days), and participate in a one-month Even Start summer enrichment program (60 ECE hours).

Program Intensity Calculations for Child Education

Preschool: (4 hrs. X 180 days) + 60 summer hours = 780 hours/year

K-2 Children: (2 hrs. X 80 days) + (3 hrs. X 32 days) + 60 summer hours = 316 hours/year

Comparing Program Intensity Totals to Recommended Hours

Preschool Education: 780 hours/year or an average of 65 hours/month (meets recommendation)

Participant attendance and attendance rate. Participant attendance data include 1) the actual number of hours attended and 2) the percentage of hours attended based on participants' possible hours (or attendance rate). Analyses of a participant's attendance and attendance rate provide the most complete picture of an individual's participation pattern. Attendance rates take into account differences in participants' enrollment dates, absences due to excused leaves from the program (e.g., pregnancy, illness), and individual participation plans to accommodate family needs and work responsibilities. These data are useful for identifying both participants who are attending at sufficient levels of intensity to achieve their goals and participants who may need additional support or motivation to improve their attendance.

Many projects communicate established participation expectations or attendance rates to families as part of the program requirements. For example, a program may specify that families are expected to participate in 75% of the hours offered in each of the four core instructional components. In addition to documenting actual hours of attendance, analyses of attendance rates allow staff to monitor family participation by component.

Participant attendance reflects the actual number of hours that individuals attended classes or participated in program services. Attendance information should be easily obtained from class attendance records; however, projects do not always maintain attendance records with the completeness and accuracy desired for Even Start program evaluations. Because incomplete or inaccurate data are useless for decision-making, and ultimately waste project resources and time,

³ Considerations in Calculating Program Intensity for Adult Education. Adult Schools of Education usually offer many classes in a subject—an Adult School may offer five levels of English as a Second Language classes and offer classes during morning, afternoon, and evening sessions. Even Start participants would be expected to attend the class that meets their literacy needs; the calculation for program intensity should take into account the number of instructional hours offered for that class. The program intensity calculation would not be the sum total of hours offered by all Adult School classes.

project staff members are encouraged to review the suggestions for quality data collection in the next section.

Calculating attendance and attendance rates. Calculating participants' attendance data consists of adding the hours attended over the course of a year. Daily attendance data should be added up monthly and then aggregated over the months that the participant is enrolled in the program during the year. To calculate a participant's attendance rate, divide the attendance hours by the total number of hours that the participant could have attended, as determined by his or her enrollment date. To look at overall participation patterns for a month, quarter, or year, evaluators should compute the median attendance and attendance rates as well as the mean or average attendance and attendance rates. Given the typically wide range in hours attended over the course of a project year, the median number of hours may provide a more accurate measure of actual participation patterns.

Example: ESL Adult Attendance for Project Year 2002-03			
Number of Total ESL Hours Offered by Program = 600 Hours			
<i>Participant Name</i>	<i>ESL Attendance</i>		
	<i># Hours Attended</i>	<i># Hours Possible to Attend</i>	<i>Attendance</i>
<i>A</i>	<i>504</i>	<i>600</i>	<i>84%</i>
<i>B</i>	<i>300</i>	<i>600</i>	<i>50%</i>
<i>C</i>	<i>300</i>	<i>400</i>	<i>75%</i>
<i>D</i>	<i>125</i>	<i>200</i>	<i>63%</i>
<i>E</i>	<i>50</i>	<i>50</i>	<i>100%</i>
<i>Median</i>	<i>300</i>	<i>400</i>	<i>75%</i>

The data in the preceding table can be used to look at individuals' participation patterns as well as group patterns by type of program service. These data identify individuals who 1) received a sufficient amount of instruction to increase their English language proficiency (see hours attended for participants A through C), and 2) participate regularly and meet the program's expected attendance rate of 75% (see attendance rate for participants A, C, E).

Duration. Family duration data describe the length of time that families remain actively enrolled in the program. Participant duration represents the number of months or years that the family actively participated in program services. Both participant attendance and duration data are important for documenting the intensity of a family's experience with program services.

Example: Attendance Rate Analyses

Mary and John both completed 100 hours of adult education. Mary has been enrolled in Even Start for six months but only attended 50% of the hours that were possible for her to attend; John has been enrolled for three months and has attended 90% of the hours possible for him to attend.

John's attendance rate indicates that he participates regularly and consistently, and is more likely than Mary to achieve his program goals in a timely manner. Further, Mary's attendance rate of 50% is well below the program expectation that participants attend 75% of their possible hours.

Project staff has scheduled a meeting with Mary to review expected participation levels, determine how staff can help her improve her attendance, and communicate that Mary's continued enrollment will depend on her willingness to participate at the expected levels.



Participant duration data are calculated by adding the number of months in a project year (or across project years if the family has been continuously enrolled for more than one year).

Managing and Analyzing Data

Participation data are useful to the extent that the data collected are complete and accurate. As stated above, attendance data should reflect all instructional services provided by the Even Start partnership and other collaborating agencies. It is anticipated that project staff will be responsible for determining the instructional services to be included in participation calculations, collecting attendance data, ensuring that the data are complete and accurate, and perhaps inputting the data into a database. It is recommended that the evaluator:

- work with staff to determine eligible instructional services;
- design a data collection system, including consent forms;
- develop attendance recording forms and provide training on form completion and data entry (if relevant);
- analyze the data; and
- share findings with project staff on an ongoing basis.

Suggested strategies for ensuring the collection of quality participation data include:

- Obtain written consent from participants indicating permission to use their data for program evaluation and progress monitoring purposes.
- Inform all program service providers of the need to collect daily attendance data on participants served.
- Provide forms for collecting attendance data and ensure that all providers of program services are trained in the protocol for completing forms.
- Require staff to submit attendance records weekly or monthly. Designate one staff member as responsible for collecting all attendance records and reviewing data to identify missing or incomplete records.
- Designate one staff member who has received training in data entry as the sole person responsible for entering all attendance data into a database or spreadsheet.
- Designate staff members as responsible for tracking the “number of possible hours” for a participant based on 1) the participant’s date of enrollment in or exit from the Even Start program and 2) the participant’s status, i.e., active or on leave.
- Work with the evaluator to develop a record-keeping system to track participants’ enrollment and exit dates. These records should also document the reasons for each family’s exit from the program and show whether the family had achieved their program goals when they left.

Analyzing participation data monthly or quarterly and providing these findings to project staff regularly will improve the project staff's effectiveness in monitoring family participation and targeting areas for program improvement. Analyses of attendance data, i.e., calculating the number of hours attended and number of hours possible to attend, and determining the percentage of hours attended, can be easily done with Excel, Lotus 1-2-3, or a similar spreadsheet program.

Using Participation Data

As stated throughout this chapter, participation data are a critical source of information for interpreting participant outcomes and guiding improvement of Even Start program services. The items below summarize how to use participation data for continuous improvement and demonstrating program effectiveness.

- **Sufficient program intensity.** Program intensity data can determine whether instructional offerings are aligned with recommended intensity levels for the core instructional components and if participants have sufficient instructional opportunities to achieve their program goals.
- **Sufficient participation to achieve outcomes.** Analyses of attendance data can identify participants who have received, or who are likely to receive, sufficient instruction to achieve their goals. When used as analytic variables, participation data can offer possible explanations for differences in participants' achievement rates and demonstrate how attendance may be related to attaining performance indicators.
- **Achieving short-term outcomes.** For many Even Start families, following through on a commitment to attend program services on a regular and consistent basis is an important first step toward achieving their long-term literacy goals. Working with families to establish attendance goals communicates the importance of regular attendance and offers an opportunity to reward and reinforce the achievement of a critical short-term outcome, that is, consistent and regular attendance.
- **Attendance rates and compliance with project attendance policies.** Many projects communicate participation requirements to families as part of the program policy and expectations for continued program participation, e.g., families are expected to attend 75% of the class hours provided in each of the four core instructional components. Individual projects should establish attendance rates that are both realistic for the families they serve and at a level that ensures enough instructional opportunity to achieve goals. Data on attendance rates let projects monitor family participation patterns and document the extent to which participants attend program services regularly and consistently.

Example: Using Program Intensity Data

Situation: *At the end of the project year, project staff members were surprised to learn that their adult participants did not meet the state performance indicator expectations, and that analyses of their outcome data showed minimal differences between adults' pre- and post-test scores. At the evaluator's request, staff carefully documented the number of hours of ESL instruction offered over the project year and the adult attendance at those classes. The data showed that the project offered about 100 class hours of ESL instruction and that adults attended classes regularly.*

Conclusion: *Although participants regularly attended their classes, most did not achieve their expected outcomes. On reviewing the U.S. Department of Education recommendations for adult education hours (see page 56), staff concluded that 100 hours of instruction a year was probably insufficient to help participants achieve their goals of increased English language proficiency. Staff members examined options for increasing the ESL class hours, such as adding more hours to the school day, increasing the number of days a week that ESL was offered, and supplementing in-class hours with distance learning.*

- **Recruiting most appropriate families for Even Start services.** Although Even Start projects are directed to serve those families “most in need,” an important “lesson learned” is that not all families can or will commit to the level of participation required to achieve Even Start program outcomes. An important use of participation data is to determine whether projects have successfully recruited families who are in most need, can attend regularly, and remain in the program for a sufficient time to achieve their goals.
- **Identifying families in need of program intervention and support.** Projects that review participation data monthly or quarterly can use these data to identify families who need more support or encouragement to increase their attendance. Sharing attendance data with a family is an effective way to show the reality of its attendance patterns and can open a discussion on the relationship between consistent attendance and reaching goals.
- **Participant perceptions of program services.** Attendance data can be used as a measure of participant satisfaction with program services. It is not uncommon for participants’ attendance rates to vary by component. Project staff may review attendance rates to see if they differ by component, which may suggest the need for further investigation. For example, analyses of attendance rates may reveal that participants attend ESL classes at an average rate of 80% but only attend parenting education at the rate of 50%. In this case, staff may wish to gather more data from parent surveys, program quality reviews, etc., to identify any issues with the parenting education component.
- **Monitoring post-test assessment schedules.** Many states have adopted performance indicators that include a requirement that post-testing occur after participants complete a specified number of instructional hours (e.g., adult education participants should be post-tested after completing 100 hours of adult education). Attendance data analyses can identify individuals who have completed the requisite instructional hours and are ready for post-testing.



Chapter 7

Learning Milestones

This chapter encourages evaluators to become involved in influencing the data that are routinely collected about learning progress, and use that information in local evaluations. It describes the following ways that evaluators can support the use of learning milestones:

- identifying milestones;
- identifying measures to assess progress;
- keeping systematic records;
- reporting progress;
- troubleshooting results;
- relating outcomes and milestones; and
- using milestone data for continuous improvement.

In addition to measuring participants' progress against state performance indicators at key times (typically once a year) family literacy programs should also follow the short-term progress of family members. Staff members need to know whether the daily instruction they are providing is helping participants reach outcomes. In order for Even Start families to stick with a program long enough to achieve outcomes that will meet family goals, Even Start families need to see that they are making progress. For both reasons, monitoring progress daily or weekly with measures aligned to learning outcomes is a key part of operating an Even Start program. (See Chapter 3 for information about measuring outcomes).

From Subpart 3—William F. Goodling Even Start Family Literacy Programs, SEC. 1235 PROGRAM ELEMENTS.

Each program assisted under this subpart shall—

(11) encourage participating families to attend regularly and to remain in the program a sufficient time to meet their program goals;...

(13) if applicable, promote the continuity of family literacy to ensure that individuals retain and improve their educational outcomes;

Family literacy staff members can use a variety of techniques and measures to follow adults and children as they acquire learning strategies and skills and develop interests and dispositions. All Even Start programs need clear benchmarks in these four areas to monitor learning progress:

- expected skills within levels of adult learning (adult basic education, English as a Second Language, study for a high school diploma or equivalent, and job readiness skills);
- language and literacy development for children ages birth through primary grades;
- other domains of development (cognitive, socio-emotional, and physical) in children ages birth through primary grades; and
- parents' stages of learning how to support children's literacy development.

Monitoring learning progress systematically guides daily classroom and home instruction:

- The early childhood educator follows a model of language development to know what she should listen for and model in her conversation with preschoolers (e.g., prepositional phrases, directional words, adverbs).
- The adult education instructor checks to see if adult students need more practice in understanding the meaning of common prefixes before moving on to a new set of skills.
- A home instructor observes whether or not a parent is ready to use dialogic reading techniques and determines how much support the parent needs to take part in a school conference successfully and obtain information to support their children's learning.

These are all examples where understanding progress milestones informs the choice of next instructional tasks.

Research or theory about the trajectory of learning and development specify milestones (as in the early childhood example above). A program may also use benchmarks built into a published curriculum (perhaps true for the adult example above). Whatever the source of milestones, they apply to all or most families even though the pace of expected progress will depend on individual family members' skills and background experiences, and the intensity of the program's interventions:

- Even though the same model of language development informs the staff's interactions with all children, the teacher's conversation with a three-year-old who has just begun to use two-word sentences will sound quite different from her conversation with a voluble three-year-old who has been in the program since infancy and who is expanding his vocabulary with adjectives for size and shape.
- An adult with only four years of formal education who is tentative about her own English speaking may take longer to achieve the milestone of helping her school age children with their homework than the native English-speaking mother who, prior to entering Even Start, simply did not know she had a role to play in helping her children with their work.

In addition to informing staff members' instructional planning, records of participants' attainments serve the important function of motivating staff and students. Acknowledging that students are achieving small steps toward longer-term outcomes helps staff recognize the benefits of their efforts—especially important when working with families who experience a number of “stops and starts.”

Monitoring progress includes:

- *reviewing the family's monthly attendance in various activities;*
- *learning whether family goals have been met;*
- *analyzing the changes in goals families set over time;*
- *tracking toddlers' language development milestones;*
- *keeping track of preschoolers' knowledge of letters and sounds;*
- *tracking reading rate/fluency of primary grade students and adults;*
- *recording numbers of books read independently;*
- *recording known English sight words or percent of correct responses to comprehension questions;*
- *keeping portfolios of writing progress;*
- *documenting the development of a family's positive relationship with children's teachers;*
- *and so forth.*



Participants also need to know that their efforts on small steps will add up to important goals. Systematically recording progress helps families see and celebrate the small steps that are associated with longer-term goals. Reading a list of common English sight words might be a small, but meaningful, milestone for the adult who is beginning to speak English. Achieving perfect attendance for the quarter might be a small milestone for the school age children of an Even Start family. When programs celebrate participants' small achievements, they contribute to the long-term retention of families in the program.

The role of the evaluator. A local Even Start evaluator might not have previously been involved in tracking participants' attainment of short-term milestones—achievements that are short of the outcomes expected for statewide program evaluation and reporting. To animate evidence-based program improvement, however, paying attention to how programs set and track progress can have a high payoff value. The evaluator can play any one or all of these six roles in monitoring progress. Each is explored briefly in this chapter.

1. Help project staff identify milestones or benchmarks associated with family goals and program outcomes;
2. Help identify tools or approaches for monitoring progress that are aligned with expected outcomes;
3. Help set up record-keeping systems to track progress efficiently;
4. Summarize and report progress on milestones across participants;
5. "Troubleshoot" failures of subgroups of participants to reach goals; and/or
6. Analyze the relationship between achieving outcomes and attaining progress to understand more about patterns of progress that predict success.

The first three roles provide consultation in assessment to program staff while the last three connect progress monitoring to other aspects of evaluation. Likely benefits of involving the local evaluator are a more systematic and objective approach to monitoring participant progress and more nuanced explanations of the processes associated with achieving (or not achieving) expected participant outcomes.

Identifying Milestones

One of the first questions an evaluator might ask program staff is, which models of skill development or curricular objectives do staff members use to plan instruction? Learning about curricular objectives and the expected sequence of meeting them gives the evaluator a fuller picture of program processes and implementation, and lets instructional staff articulate their understanding of the learning process. (See similar advice about assessing program implementation in Chapter 5.)

If a new project does not have curricular objectives in all domains or does not suggest a sequence of development, the evaluator of a first-year project might suggest locating an appropriate framework to guide instruction, and help select a framework that is in line with expected program outcomes. For example, if the performance measure for preschool

children is a language assessment that stresses vocabulary and comprehension, the program should have some language development milestones for young children that are related to understanding the meaning of high-frequency words in various categories, e.g., words for position, order, scale, senses.

Sources of curriculum milestones include:

- state content and performance standards for PreK-12;
- local district content or curriculum frameworks;
- the theoretical model underlying commercial curricula or assessments; and
- syntheses of research such as *Preventing Reading Difficulties in Young Children* (National Research Council, 1998) and the *Framework for Parenting Education* (U.S. Department of Education, 2000) which summarize information about how children learn to read and parents' roles in supporting children's learning.

Family goals. Many Even Start programs ask families to identify the goals that they expect to achieve through participation in family literacy. Goals are often a mix of learning objectives, e.g., obtain a GED, learn to speak and read English, or earn nurse's aide certificate, etc., and improvements in family functioning, e.g., get a better apartment, obtain a driver's license and car, or become a citizen. Family goals can serve as milestones for tracking a family's sense of efficacy and purpose—critical for retaining families in the program long enough to meet learning goals.

Setting goals is sometimes part of recruitment and orientation activities. Family goals provide insights that staff can use to frame instructional tasks; for example, the driver's preparation manual or church bulletin may become informational text for lessons about word patterns. Some programs use goal-setting to introduce families whose lives may be otherwise disorganized and crisis-oriented to planning for long-term change. Staff can help guide goal-setting by breaking large ambitions into smaller achievable steps and coaching adults to see connections among a series of short-term goals.

Program staff members work with families regularly to revisit their goals, assess progress, and set new goals. An alternate approach asks parents to respond periodically to a series of generic goal statements (e.g., improve housing, speak and write English, manage children's behavior, help children with homework) by noting whether or not the statement represents a current goal, an achievement, or an area that is not of concern.

Adults entering programs commonly set goals that may be unrealistic—the adult with little formal education states he is entering a family literacy program to become a doctor. Six months later, the

Example: Recommending a Framework for Parent Education

Sandi Miles is the evaluator of LEARN MORE, a new Even Start project that collaborates with the local community college and the school system. Miles quickly learns that the adult basic education staff follow a structured curriculum that incorporates word study practice with comprehension strategies. For every eight lessons, the staff administer an assessment of mastery of elements of word structure, fluency, and comprehension. But the program has no similar structure for parent education. Each family educator independently plans home lessons to encourage parents to work with their children on language and literacy activities. In her first-year interim evaluation report, Miles recommends that the project develop a set of research-based goals in parenting education as an umbrella for lesson development and progress monitoring. Miles recommends that staff contact the local university-based parent education institute for guidance. Since one outcome for preschool children is story comprehension, Miles suggests that curriculum objectives for parents at least include mastering the principles of dialogic reading for shared conversations about books with their children.

Example: Recommending a Progress Measure

In its second year, project LEARN MORE has implemented the Pearson Education dialogic reading program with parents and has set other goals for parents in talking with children and setting behavior limits. Staff members are eager to assess growth in the parent role because they feel they are starting to see real differences in some parents and would like tangible evidence of their work.

The evaluator introduces the staff to the Parent Reading Belief Inventory, an instrument that taps into areas related to parent-child joint reading, and the Parent Education Profile (PEP) (RMC Research and NY State Department of Education, 2003), a comprehensive framework that organizes parent's behaviors along progress continua, including an Interactive Literacy scale. The evaluator suggests piloting each instrument with a few families to determine their utility and practicality for the LEARN MORE project.



same student's stated goal is to earn a GED and obtain a driver's license. Simply tracking the progress toward goals may not be illuminating for an evaluator, although following the changes in the nature of family goals over time could provide insight into an adult's ability to set and achieve meaningful and realistic goals. The evaluator can also use family goals as a backdrop against which to assess the appropriateness of a program's instructional activities or to interview parents about their satisfaction with their rate of learning progress.

Identifying Measures to Assess Progress

Program directors are likely to call upon evaluators' expertise to help select, evaluate, or create progress measures. The caution here, of course, is that developing progress measures can quickly grow beyond the scope of a typical evaluation.

The local evaluator brings knowledge of the outcome measures used for performance indicators and the judgment to help project staff assess the alignment among possible measures of milestones, curricular goals, and outcomes. Is the proposed progress measure valid—does it give a “reading” on the most important milestones? For example, does a proposed observation record of child progress provide enough of a “reading” on language development to indicate how well children will do on the program's end-of-year assessment? If the observation record largely focuses on physical and social development, it will provide useful information but not be as salient as a recording form that also prompts staff to observe language use.

Often projects assess learning progress with locally-developed systems. Evaluators can help staff members figure out how to use the information that they routinely collect—in portfolios, for example—to determine whether or not participants' progress is adequate. Evaluators might help staff develop a list of possible portfolio samples and locate scoring rubrics for judging degrees of skill mastery. Portfolio items could include: a journal entry to assess skills in written English, a tape recording of the student reading informational text to judge reading rate and fluency, and examples of student work to assess transfer of job readiness skills.

The local evaluator can also help program staff learn about, gain access to, and use the results of the progress measures primary grade teachers use to track the development of literacy skills. Schools increasingly monitor early reading progress systematically. Familiarity with various assessment tools, e.g., the Dynamic Indicators of Basic Early Literacy (DIBELS) and fluency monitors, can guide the program staff in offering supplementary instruction during home visits and in after-school programs.

Keeping Systematic Records

An ideal evaluation task for a first-year project is ensuring that a project has solid record-keeping systems that all staff use. It can be much more challenging to consult with a mature project where individual instructors have implemented a number of separate systems. In some cases, the evaluator's most helpful role may be leading staff to organize the information they already routinely collect about participant progress.

First, the evaluator might guide the staff in creating an inventory of the milestones and measures that staff members already use and determine which are essential for all families. These might be milestones used by other agencies that provide instruction, such as Head Start and adult education programs, K-3 teachers, and social service support agencies.

Information on milestones that apply to all families forms the basis for a family progress record. The evaluator might facilitate record development and also advise on computerizing and/or linking progress records to existing software systems to record performance indicator information for state-level reporting. A relational database is ideal because progress information for individual family members can be organized by family units and also summarized across groups of individuals. This will aid in reviewing results and planning instruction. The evaluator might consult with project staff as they design such a database or help locate appropriate technical expertise.

With information organized efficiently, the evaluator can help project staff develop and use data summaries to seek patterns that apply to groups of participants. Visual displays of information are especially useful in discerning patterns of progress or gaps in progress.

Example: Looking at Progress Data

Jorge Perez has organized AYUDA's data about adult basic education unit tests in several ways to aid the staff's review of progress. This chart inspired staff to reexamine when they scheduled assessments and what motivational incentives they might offer long-term and lower-functioning students.

Groupings of Students	% correct medians	% correct range	Avg. length of time in program
<i>Students who entered below grade</i>	<i>35%</i>	<i>8%-68%</i>	<i>13 months</i>
<i>Students who entered at grade 8 and above</i>	<i>70%</i>	<i>55-100%</i>	<i>10 months</i>
<i>Students who attend less than 50%</i>	<i>42%</i>	<i>10-75%</i>	<i>16 months</i>
<i>Students who attend between 51-75%</i>	<i>85%</i>	<i>70-95%</i>	<i>14 months</i>
<i>Students who attend more than 75%</i>	<i>86%</i>	<i>68-100%</i>	<i>12 months</i>

Example: Sandi Miles Reports on Progress

Miles is able to report some information in the LEARN MORE project's third-year evaluation report about the specific progress of Even Start parents.

After at least one full year of Even Start participation, documentation shows about 60% (increased from about 40%) of parents:

- *Demonstrate awareness that their language affects their child's language and behavior and try strategies to support the child's development;*
- *Show interest in learning how to tell stories and read to children and can use a few strategies to engage children; and*
- *Have begun to help children learn how print works.*



Reporting Progress

The benefit of involving the evaluator in designing a program's record-keeping system is the familiarity he or she will develop with data sources, which can inform other aspects of the local evaluation. Not all local evaluations will necessarily use progress monitoring records—that will depend on the evaluation question guiding the local evaluation plan. Below are examples of focused inquiries in which the progress information staff members routinely collect would be an essential data source:

- *Are the preschool children served by the child care center making as much progress in language development as the children in Head Start?*
- *How much progress do adults who drop out before taking the post TABE (the measure used for performance indicators) make?*
- *How much more progress do families who receive two home visits per week make compared to families who receive one visit a month?*

For those inquiries, tracking and summarizing progress is essential in responding to the evaluation question. In other cases, progress monitoring data would be less central to the evaluation but might provide additional explanation of other results. For example, the evaluator might summarize learning progress in more detail to illustrate the meaning of an average one level gain on a pre-post assessment.

Troubleshooting Results

When outcomes are disappointing, the evaluator needs to explore the reasons why in order to make viable suggestions for improving a program. In addition to analyzing items on the outcome measures, checking participation levels, and determining quality of instruction, tracking what is known about patterns of progress should be central. For example,

- *If a preschooler is at the same percentile level on the Preschool Language Scale after a year in Even Start, what does her portfolio show about her progress in using categories of vocabulary words, the complexity of her sentence and grammatical structures, and her ability to respond to questions about stories?*
- *If an adult basic learner did not make a grade level gain on his TABE after a year in the program, what do end-of-unit tests and informal assessments suggest about his reading progress?*

Answers to such questions are important clues to whether or not the results of the outcome measure accurately reflect participants' progress, whether there is cause for concern about the overall pace of progress or the intensity of the intervention, and/or whether the progress milestones that inform instruction are the right guideposts.

Relating Outcomes and Milestones

Given that Even Start programs typically serve fewer than 50 families, it can take a few years to amass enough data about subgroups to see patterns. The evaluator of a mature program may have accumulated enough information to look for patterns in learning progress that predict performance on outcome measures. In the spirit of motivating staff and parents to understand more about how their steps toward progress can add up to important goals, evaluators could analyze the relationship between outcomes achieved on performance indicators and progress made on learning milestones. Examples of evaluation questions that could guide inquiries are:

- *What language skills did the four-year-olds who scored in the highest category of the Get Ready to Read assessment demonstrate during quarterly language reviews?*
- *What is the average score on a practice test that predicts success with the GED?*

Answers to focused inquiries such as these have many benefits. They guide staff in decisions about program design and instruction and give parents realistic guideposts for gauging the time and effort required for results.

Using Milestone Data for Continuous Improvement

Data on milestones that participants have or have not achieved are important for discussions about program improvement. Some ways to use such data for improvement follow.

- **Feedback on instruction.** The evaluator can summarize the types of milestones families achieved or did not achieve and how long it typically takes to achieve key milestones. The resulting information can help show whether program services are intense enough, practice opportunities are adequate, and goals are realistic.
- **Differences by site and instructor.** Depending on how the program is structured, it may be possible to analyze milestone information by instructor or site to learn if some effective learning practices might be replicated in other settings. For example, the students of one adult education instructor may consistently achieve higher results on unit tests. The evaluator might conduct a further inquiry into the practices and strategies that contribute to

Example: Understanding Progress
John Tabor was understandably disappointed with the primary grade results this year from the HILL Central Even Start program. When he checked school records, he found that only one of the ten children from Even Start families was rated as reading at or above grade level for the year even though the program had finally begun after-school supplementary tutoring in reading—a recommendation he had made in last year's evaluation report.

John decided to see what he could learn from the other nine children's teachers about their progress. Checking the quarterly early literacy profiles the school completed, he found that seven of the nine children had made appropriate progress for their grade levels in decoding skills, and that all nine recognized common sight words. Only one of the nine was rated as adequately fluent in oral reading; all nine were struggling with even basic comprehension details such as retelling story events and identifying main characters. As a result, the nine children were two to three levels behind most children in their classrooms.

Instead of simply reporting that only 10% of HILL Central's children were successful in reading at or above grade level (the state's performance indicator), John was able to supplement the information with this progress chart.

Percent Achieving Adequate Progress in:

<i>overall reading level</i>	<i>10%</i>
<i>decoding skills</i>	<i>70%</i>
<i>sight word recognition</i>	<i>100%</i>
<i>fluency</i>	<i>20%</i>
<i>literal comprehension</i>	<i>10%</i>

Example: Predicting Success

Don James carried out a longitudinal study to track four cohorts of Even Start children who had entered school, about 100 children. From the data on reading progress, he determined that about 65% of the students read on grade level or above and were deemed by their teachers to have solid foundations of reading skills.

For the purpose of developing recommendations, James is curious about those students' profiles when they were in the Even Start pre-K program. Using the project's computerized record system, he will review the results of skill mastery checks that teachers carried out every six weeks. He wants to explore some specific potential predictors:

- *When did the students with reading competence master initial and final letter sounds?*
- *Did they enter kindergarten able to read sight words?*
- *What do we know about their vocabulary development at ages three and four?*

the strong showing, and document them so that program leaders can improve the pedagogy other staff members use.

- **Differences by groups of participants.** Examining patterns of progress may help program staff identify characteristics of participants that point to different instructional needs. For example, some adult participants may have more difficulty writing sentences and paragraphs than others. Further study would look at the path of learning progress in other goal areas: does this group tend to progress slowly with comprehension or vocabulary? Does it participate as actively in the program, etc.? The answers could inform how staff groups learners for instruction.
- **Reflecting on outcomes.** When annual outcome results are disappointing for a whole group, it is worthwhile for the evaluator to analyze the results in light of progress on related milestones. This may suggest areas where materials or pedagogy should be strengthened. For example, if analysis of children's tests shows systematic weaknesses in understanding vocabulary words, a next step would be to learn whether or not vocabulary development is tracked and, if not, to develop a plan to do so. If vocabulary development is tracked, the next step would be to examine the progress monitoring results. They may show that vocabulary development is solid and instruction is adequate, or they might show weak or average development that needs enrichment through more opportunities for conversation and play.
- **Critical points.** Analyzing progress patterns in relationship to program drop-out patterns may illuminate the places where participants "get stuck" or become discouraged about their ability to make progress. With this information, staff can restructure program offerings, provide extra attention to participants at key points, identify incentives for effort and progress, locate additional learning aids, and/or organize peer support.



Final note to program staff members. This chapter is filled with ideas that may extend well beyond the scope of current local evaluation expectations in some states. In the event that all the roles described in this chapter are new, the options may seem overwhelming. Program staff members should begin by considering which might have the greatest payoff for improving instruction in the local program. If the program expects the evaluator to serve as a consultant and help identify or develop learning progress measures, the time involved could be substantial and would likely require additional compensation for the evaluator.



Chapter 8

**Selecting an
Independent Evaluator**

This chapter covers selecting an independent evaluator. It offers guidelines to ensure that the evaluator/staff partnership is a productive working relationship. The suggestions outlined in this chapter are not prescriptive; projects should select the options that meet their needs and are feasible to implement given available resources. Topics in this chapter include:

- defining independent evaluation;
- desired qualifications of Even Start evaluators;
- setting priorities for the local evaluation and evaluator; and
- selecting an independent evaluator.

Defining Independent Evaluation

The program statute, Section 1235 (15) of the ESEA, requires Even Start projects to provide for an independent evaluation of their program that they will use for program improvement. An independent evaluation is conducted by one or more individuals who are not directly involved in the project's administration. That is, projects must use an evaluator who is outside of or external to the Even Start project. Such evaluators are often referred to as independent, outside, or external evaluators.

In an independent evaluation, the Even Start project director, in consultation with the evaluator, determines the evaluation's purpose and overall scope. The evaluator will plan and conduct the study and provide an independent report of the findings. The rationale for employing an independent evaluator is to ensure that projects benefit from the perspective of an objective and unbiased "outsider."

Independent evaluators usually work as consultants to the Even Start program and generally are not employees of the partners responsible for administering the program. An exception to this rule, for example, would be large school districts or institutions of higher education that work in partnership with an Even Start project. In such cases, an employee of the institution may serve as the project evaluator provided that he or she does not work in a division involved in administering the Even Start program or the department responsible for providing program services.

Desired Qualifications of Even Start Evaluators

The Standards and Guidelines adopted by the American Evaluation Association (AEA) are useful in beginning a discussion about the desired qualifications of professional evaluators. These Guidelines set standards of professional practice and inform clients about the principles they can expect professional evaluators to uphold.

AEA Guiding Principles for Evaluators⁴

Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

Evaluators provide competent performance to stakeholders.

Evaluators ensure the honesty and integrity of the entire evaluation process.

Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.

Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

The AEA principles emphasize qualities such as the ability to conduct systematic research, competency, honesty, integrity, and the evaluator's respect and sensitivity for all involved in the evaluation process. The specific qualifications of Even Start evaluators proposed in the following section reflect the AEA principles that are likely to be relevant to most Even Start program evaluations. The examples show evaluator qualities considered less helpful for achieving the goals of Even Start evaluations.

Desired Qualification # 1: Evaluators possess evaluation and measurement expertise.

Independent evaluators who conduct Even Start evaluations should possess the education, skills, and experience needed to design and conduct data-based evaluation studies. The evaluator should have formal training in evaluation design and measurement as well as experience applying those skills in evaluations of educational programs. It would be most helpful to Even Start projects if evaluators possessed the technical expertise to:

- develop evaluation designs with meaningful evaluation questions;
- evaluate the technical qualities of assessment instruments and help staff select appropriate instruments;
- develop surveys, interview protocols, or other desired instruments;
- monitor the collection and management of evaluation data;
- analyze quantitative and qualitative data;
- write evaluation reports and communicate findings to diverse audiences; and
- translate findings into specific program recommendations.

Desired Qualification # 2: Evaluators are knowledgeable about family literacy programs and practices.

Even Start family literacy programs are complex and challenging to implement. Evaluators who understand the mission and practices of family literacy programs can support staff by helping to define key evaluation questions and interpret findings in light of common family literacy issues. Ideally, projects should seek evaluators with experience in evaluating family literacy programs. Since this expertise may not always be available, the next best strategy would be to

⁴ For additional information on Guiding Principles for Evaluators, see *A Report from the AEA Task Force on Guiding Principles for Evaluators* (1999).

Example: Expertise

A retired school principal, Mr. A. was instrumental in getting the Even Start program funded and is excited about the possibility of continuing his involvement by serving as the project evaluator. He is knowledgeable about family literacy and figures he can learn about evaluation "on-the-job."

Selection Issues: Without a background and some experience with program evaluation, Mr. A. can only provide limited guidance on evaluation issues. It is unlikely that he will be able to perform the full range of evaluation tasks expected of an independent evaluator.

Example: Independence

A school district employee, Mr. B. teaches computer science in the Adult Education division of the school district awarded an Even Start project. He has a background in data analysis and was approached by the Adult School Director, who is also the Even Start project director, about evaluating the Even Start program at their site.

Selection Issues: While his limited background in program evaluation may also be an issue, Mr. B. does not qualify as an "independent evaluator" because he works for the division of the school district that administers the Even Start project.

Example: Independence

Dr. C. is a professor at the university. She is known for her research in parenting education and has experience in program evaluation. The Even Start project has hired her to provide training, monitor implementation of the parenting curriculum that she developed, and evaluate the Even Start program.

Selection Issues: *Dr. C. possesses evaluation expertise but her ability to function as an independent evaluator is compromised because she provides services as a content expert and will evaluate implementation of her parenting curriculum.*

Example: Working Style

Ms. E. has years of experience working as an independent evaluator. She provides the project with an evaluation design and a binder of materials listing all instruments and a data collection plan. She tells the staff that her role will be to monitor their data collection efforts over the course of the year and write the report.

Selection Issues: *Ms. E.'s directive style does not facilitate staff and stakeholder collaboration in the evaluation process. The lack of staff involvement may mean that the project's information needs are not addressed in the evaluation findings.*

select evaluators who are willing to invest time in learning about Even Start program elements and practices.

Even Start projects sometimes consider selecting evaluators who are retired Even Start directors or consultants with expertise in a content area such as early childhood development or adult education. Although this is appealing because the evaluator brings family literacy or content expertise, evaluators should not be selected solely on the basis of their content expertise. Projects are better served by hiring experts in family literacy or a content area to help with program development, and then selecting a program evaluator with evaluation expertise. This separation of responsibilities also ensures that the content expert does not provide direct services or advice, only to be later asked to “objectively” evaluate the program.

Desired Qualification # 3: Evaluators work in partnership with staff and other

stakeholders. Stakeholders, project staff, and evaluators all play roles in designing and conducting an evaluation. Evaluation is a process based on teamwork. It requires everyone who is involved in developing and evaluating a program to work together to achieve their goals. The precise nature of the collaboration will vary by program, but at a minimum, staff and stakeholders should be directly involved in developing evaluation questions that address the project's information needs, discussing approaches for answering those questions, and reviewing data to interpret the findings. Evaluators who work collaboratively tend to seek opportunities for involving staff and other stakeholders in the evaluation; they encourage staff to express concerns and interests, value their input, and respond by being flexible and adjusting evaluation strategies to meet the needs of the group.

Desired Qualification # 4: Evaluators possess good communication and listening skills.

Evaluation expertise is only useful if evaluators can communicate in a manner that allows others to understand and interpret their work. Most project staff members have limited evaluation experience and are not familiar with evaluation terminology. Evaluators who use evaluation jargon without defining the concepts tend to inhibit rather than facilitate staff understanding; this may alienate rather than encourage staff participation in the evaluation process. To communicate effectively about evaluation, the evaluator must be able to define terms and explain concepts in everyday language that a layperson can easily understand.

As stated above, evaluation is a collaborative effort. The evaluator's role is to facilitate the involvement of others in that effort. In order to work effectively with policy makers, staff and families, evaluators must be able to communicate on a variety of different levels to different audiences. Effective communication skills also include the ability to listen and a willingness to respond to the concerns and issues of others.

Desired Qualification # 5: Evaluators demonstrate sensitivity and respect for people.

Evaluation is highly interactive. Its success depends on the quality of relationships between the evaluator and those involved in the project. To a large degree, this depends on the evaluator's attitudes and behaviors. Positive behaviors include 1) showing respect for others, 2) abiding by professional ethics (e.g., informed consent, confidentiality), and 3) communicating in a way that acknowledges the dignity and self-worth of those involved (e.g., solicits input from all staff, recognizes that successful programs are built on the strengths of the families and staff). Positive behaviors convey that the evaluator is aware of and sensitive to differences among participants (e.g., based on culture, religion, gender, ethnicity, etc.) and is mindful of the potential implications of these differences when planning and conducting evaluation studies and interacting with program participants.

Desired Qualification # 6: Evaluators demonstrate professional integrity and honesty.

Another important consideration is the evaluator's reputation for professional integrity in terms of completing tasks in a cost-effective and timely manner. Projects are encouraged to request information from the evaluator's previous employers and to ask prospective evaluators to disclose other roles, relationships, and time commitments that may affect their ability to perform the scope of work. Given that the evaluator will be entrusted with data about the program and the families served, it is also vital to consider the evaluator's reputation for honoring agreements about using data and findings. For example, evaluators should always receive specific permission from a project and families before using their data for independent research or conference presentations.

Setting Priorities for the Local Evaluation and Evaluator

Identifying information needs and evaluation requirements. In selecting an evaluator, staff members should begin by thinking about the tasks and services that they will need an evaluator to perform. The first step is to identify specific state evaluation requirements and other evaluation targets that apply to Even Start programs. Some states may also specify certain instruments and timelines for data collection and reports of project outcomes. Projects that receive funding or services through other organizations may also be required to address their evaluation requirements.

The next step is to engage staff and other stakeholders in preliminary discussions to identify key questions or information needs specific to the program. For example, a project director may be interested in surveying families about their perceptions of program services in order to determine why a high percentage of families drop out of the program after fewer than two months. He or she might ask an evaluator to enhance the evaluation by adding a parent survey to the required tasks of

Example: Communication

Dr. D. is a specialist in educational research. During his last presentation, the staff sat in confused silence as he presented results for the adult education indicators by stating that he used a crossed design and the results indicate that the null hypothesis could not be rejected.

Selection Issues: *Dr. D.'s research jargon does not provide sufficient explanation to communicate evaluation concepts. Staff members who do not understand the findings will be less likely to use the data for program improvement.*

Example: Respect

Dr. F. is an evaluation consultant. In her first meeting with the director she asks if she will be required to drive to certain parts of town to observe classes and meet with staff. She refers to the families as interesting research subjects and leaves a three-page informed consent form for the director to review.

Selection Issues: *Dr. F.'s behaviors suggest that she may not have the level of sensitivity or awareness needed to engage Even Start staff or families in the evaluation process.*



Example: Competing Tasks

Mrs. G. has a very successful grant-writing and evaluation consulting business. Although she has an excellent reputation, her recent successes and increase in business have impaired her ability to provide her clients with the same quality, customized services, such as conducting on-site visits to observe the program in operation, as she has in the past.

Selection Issues: Projects may wish to gather more information about Mrs. G.'s availability to perform additional evaluation studies.



documenting participant outcomes. (See Chapters 10 for ideas on determining evaluation questions and information needs.)

Prioritizing program needs and evaluation tasks. The initial steps should also include prioritizing the desired evaluation tasks in order to make the best use of the evaluator's time and expertise. Priority tasks would include the project's evaluation requirements and those considered essential for program improvement. Another way of prioritizing tasks is to identify tasks that are best completed by the independent evaluator and those that other program staff can accomplish. By prioritizing evaluation needs, projects will develop a more realistic picture of the services they need the evaluator to perform.

Determining levels of evaluator involvement. Once staff members have developed a working list of priority tasks for the evaluator, they should discuss the level of involvement they expect of their evaluator, both in the program and the evaluation. To a large extent, the level of the evaluator's involvement depends on the financial resources available for the evaluation, the evaluation requirements, the time available from the evaluator, and program staff's evaluation expertise. The closer the partnership desired, the greater the evaluator's time commitment and the greater the cost.

The relationship between program staff and evaluator will depend on individual circumstances and is likely to evolve over time. By thinking about the following four questions, projects can begin to determine the degree and extent of evaluator involvement they wish.

1. Which evaluation tasks should the evaluator perform (e.g., select outcome assessment instruments, write the evaluation report)?
2. Which evaluation tasks are most appropriate for others to perform, with the evaluator acting as an advisor or monitor (e.g., monitor staff data collection activities, provide advice on interpreting attendance data)?
3. What are the program's expectations for evaluator involvement in ongoing activities (e.g., conduct monthly site visits, attend occasional staff meetings)?
4. What other activities are expected of the evaluator (e.g., attend state and/or federal evaluator meetings)?

Most Even Start projects use an independent evaluator to complete tasks for which staff lacks the expertise or time, such as developing surveys or analyzing and presenting data associated with participant outcomes. Depending on the project's needs and available resources, the evaluator may be a frequent presence in project activities or be more selectively involved in key evaluation tasks.

Selecting an Independent Evaluator

Selecting the right evaluator should be considered a long-term investment. Time spent on the selection process will pay off in the long run. Optimally, projects will seek an evaluator who will continue to work with the project and provide evaluation services for the duration of the project's funding cycle. Spending time at the outset to ensure that the selection process produces the best project–evaluator match is an investment in the future and will save time, money, and stress.

The following steps for choosing an independent evaluator are guidelines; projects should adapt them to meet their own circumstances. Projects should use their own procurement procedures, reflecting applicable state and local laws and regulations, provided that the procurements conform to applicable federal law and the standards identified in this section. For example, projects that contract with an individual or business to provide an independent evaluation must follow the procedures in the applicable federal procurement standards located in the Education Department General Administrative Regulations (EDGAR).⁵ The sequence of steps in selecting an evaluator is outlined below.

Step 1: Describe program evaluation needs. It is a good idea to begin by developing a written description of the high priority evaluation services. The description might be formalized into a job description or as simple as a list of talking points for conversations with prospective evaluators. It should include:

- a brief description of the Even Start program and its scope, e.g., location, number of sites and families served;
- a statement of the purposes of the independent evaluation, e.g., to evaluate families' literacy outcomes based on the state's performance indicators, to gather participant perceptions of program services;
- a list of key tasks and services the evaluator is expected to perform e.g., to work in collaboration with staff to design and implement the evaluation, interview a sample of families, aggregate data and analyze data, write an executive summary and technical evaluation report, develop presentation for Board of Education;
- an estimate of days, e.g., requires an evaluator for approximately 25-35 days in the next year, or the specific amount for the evaluation agreement, e.g., evaluation agreement is for \$10,000;

⁵ For more information on the applicable federal procurement standards in the Education Department General Administration Regulations, see EDGAR, Sections 74.40 – 74.48 (34 C.F.R. sections 74.40 - 74.48) for institutions of higher education and non-profit organizations and Section 80.36 (34 C.F.R. section 80.36) for state governments such as state educational agencies, local governments such as local educational agencies, and Indian tribes and tribal organizations. EDGAR is available at: <http://www.ed.gov/policy/fund/reg/edgarReg/edgar.html> .

Example: Sample List of Prioritized Evaluation Tasks

Higher Priority Tasks

- *Develop evaluation plan and key questions*
- *Analyze outcome data for state performance indicators*
- *Develop parent survey*
- *Write executive summary and technical report*
- *Present findings to collaborators*

Lower Priority Tasks

- *Conduct self study with staff*
- *Aggregate parent survey data*
- *Write up findings from parent survey*
- *Attend staff meetings*



- a request for a mode of response, e.g., written proposal or meeting; and
- project contact information.

Clearly communicated expectations set the groundwork for a good working relationship and save time in the selection process. A clearly written description helps prospective evaluators determine if they are interested in the priority evaluation tasks and the expected time commitment.

Step 2: Sources for locating evaluators. The search for an evaluator should begin by checking the selection and contracting requirements of the partnership agency responsible for administering the Even Start program. The federal procurement standards, state, and/or local organization practices will govern approved sources, award preferences for evaluators from affiliated agencies, procedures for seeking bids, daily costs, and/or selection criteria.

Evaluators work in several ways. They can be independent, self-employed consultants, they may work for a company that specializes in evaluation and research, or they may be affiliated with an agency or university. Well-qualified and affordable evaluators may be found in all these configurations. Even Start projects seeking an independent evaluator might contact the following local sources:

- college and university departments that offer programs in evaluation and educational research or Even Start-related services, e.g., early childhood development;
- state agencies, including the SEA Even Start coordinator and departments of education and human services. State agencies often keep bid lists of approved evaluators;
- regional or state affiliates of related professional organizations, such as the American Educational Research Association or American Evaluation Association. Some associations now have online directories;
- other Even Start or family literacy programs willing to share their experiences and recommend independent evaluators; and
- collaborators who may have employed evaluators.

Projects in larger cities might also consider placing an open-call newspaper advertisement for evaluator services. Projects in rural areas are more likely to conduct a wider search for evaluation candidates and may need to offer incentives such as reimbursed travel expenses to attract well-qualified evaluators.

Step 3: Review responses. Prospective evaluation candidates should provide a resume and a brief written response outlining their evaluation approaches and costs (unless the amount has been pre-specified). The evaluator's proposed activities should agree with the level of involvement desired by the staff and directly address the project's evaluation needs and priorities. Project directors should be wary of "canned" approaches that do not read as though they are tailored to the program's needs. When working with an evaluation firm, a large evaluation team, or a university, the project director should identify a contact person and request the names of the evaluators who will actually perform the work.

Step 4: Assess evaluator qualifications. Projects should seek evaluators with qualifications for the specific needs of the Even Start program. Prospective evaluators' written and oral responses will help project leaders judge the candidates' qualifications.

Potential candidates should show evidence of their skills and experience by providing a resume, a summary of related work, and references from past clients. Evaluator resumes may present education credentials or degrees in research or related fields; however, in order to determine whether the evaluator can apply this knowledge and produce useful products, the staff needs to review work samples, including written reports, evaluation designs, and instruments developed for similar projects, etc. Projects are encouraged to review writing samples or oral responses using the following criteria:

- messages and findings are clearly communicated and easily understood;
- the complexity of the material is communicated in an appropriate manner for the target audience, e.g. families, staff, and other stakeholders;
- the use of technical jargon is minimal; evaluators use laypersons' language to communicate evaluation concepts;
- recommendations for program improvement are based on the evaluation findings;
- there is evidence that the evaluator's work has been used by past clients; for example, evaluation results helped to reshape program plans; and
- the evaluator offers strategies for involving others in the evaluation process; for example, his or her management plan includes staff meetings to discuss evaluation questions.

Step 5: Make the final selection. It is difficult to give specific advice about the final selection of an evaluation candidate. Finding the best candidate requires the willingness to invest the time to ensure that the selection process is as thoughtful and comprehensive as possible. Applicable federal procurement standards and/or practices in the partnership agencies or organizations may dictate final selection procedures; for example, the evaluation agreement may be awarded to the lowest bid or give preference to women or minority-owned businesses. Projects are encouraged to be realistic about the selection process, and to realize that some compromise is inevitable. Given that projects may not find the "one" evaluation candidate who embodies all the desired qualifications, the best choice would be to select the candidate with the strongest professional references based on past evaluation work.



Chapter 9

**Working With
an Evaluator**

Once projects have selected an evaluation consultant, the next step is to develop a formal agreement for evaluation services. This chapter covers:

- basic elements of an evaluation agreement;
- developing a Scope of Work; and
- the evaluator–staff relationship.

Basic Elements of an Evaluation Agreement

It is common practice to develop an agreement that covers a project year or phase of the evaluation and to update agreements as needed in later years. Even Start projects that use a contract vehicle to retain an independent evaluator should check the federal procurement standards applicable to their organization based on the Education Department Administrative Legislation Regulations (EDGAR).⁶

Agreements for independent evaluation services often include these elements:

- a scope of work that summarizes the evaluation tasks that the evaluator has agreed to perform and a timeline for completing tasks and deliverables;
- a list of deliverables or expected products e.g., instruments, formal reports, with due dates for completion;
- a list of products or services that the project will provide the evaluator, e.g., test records, attendance files, scheduling parent focus groups;
- a statement of mutual understanding about the confidentiality of records and reports, including restrictions on publishing articles or reports about the project;
- procedures for handling disputes;
- billing procedures and a schedule of payments; and
- a process for amending the agreement.

Developing a Scope of Work

The heart of the evaluation agreement, the Scope of Work specifies all the tasks and functions that the evaluator is expected to perform. The more clearly defined the Scope of Work, the higher the likelihood that the desired tasks will be completed and that expectations for the evaluation will be met. The sample Scope of Work on page 86 lists key evaluation

⁶ For more information on the applicable federal procurement standards in the Education Department General Administration Regulations, see EDGAR, Sections 74.40 – 74.48 (34 C.F.R. sections 74.40 - 74.48) for institutions of higher education and non-profit organizations and Section 80.36 (34 C.F.R. section 80.36) for state governments such as state educational agencies, local governments, such as local educational agencies, and Indian tribes and tribal organizations. EDGAR is available at: <http://www.ed.gov/policy/fund/reg/edgarReg/edgar.html>.

tasks that could be included in a comprehensive Even Start local evaluation. Some tasks are broken down to show more detailed examples of potential services, products, or activities.

For each key evaluation task, the Scope of Work estimates the time required to complete the task. These estimates are described in terms of days; the evaluator's daily rate may determine the actual cost for each service or product. Although evaluator rates are likely to vary widely by state and city, the daily rates of experienced evaluators may be about \$500/day. Project staff should confer with their state coordinator or other Even Start programs about reasonable daily rates for evaluator services in their area.

Remember that if evaluators join the project as independent consultants, their daily rates will likely include all overhead-related costs (e.g., office rental, office equipment, secretarial, printing, telephone, payroll taxes, and liability insurance). Travel expenses are often not included in an evaluator's daily rate, and if an evaluator is required to travel considerable distances to the project or attend required state or federal required meetings, this expense may be an added negotiated cost (to the extent allowable).

The actual number of days needed to complete tasks is subject to discussion and negotiation with the independent evaluator. It is important to note that the number of days suggested in the sample Scope of Work are "time estimates." The actual tasks in a project's Scope of Work may require less or more time. It is also important to remember that the "total number of days" in a Scope of Work generally represents 1) the days that the evaluator works on-site (e.g., conducting site visits, interviewing parents, participating in staff meetings), 2) days he or she works off-site (e.g., analyzing data, writing reports), and 3) days the evaluator attends required state and/or federal trainings. For example, it would not be unusual for an evaluation agreement of 30 days to include 12 days of on-site time, 17 days of off-site time, and one day for a required training.

The following Scope of Work, provided solely as a resource, incorporates the key evaluation tasks proposed in each section of the *Guide's* evaluation framework (Chapters 3 through 7). Conducting a comprehensive local evaluation designed to yield useful information for staff and other stakeholders requires the allocation of sufficient resources. Projects that choose to implement the recommendations outlined in this *Guide* may need to consider budgeting between \$10,000 - \$15,000/year for their local evaluation.

Sample Scope of Work for a Comprehensive Evaluation of an Even Start Program
(Includes approximate number/range of days for completing each task)

Tasks/Activities/Products	Days @ \$500/day
1. <i>Develop an evaluation design and management plan</i> a. Review state evaluation requirements, performance indicators, and other ES statutory requirements b. Meet with staff and other stakeholders to identify participant outcomes, evaluation focus, and key evaluation questions c. Select and/or identify assessment instruments d. Develop a data collection plan e. Develop a timeline and management plan for conducting study	3-4 days
2. <i>Train staff in administering, scoring, and interpreting selected instruments</i>	1-2 days
3. <i>Develop forms, interview protocols, and questionnaires</i> a. Consent form, attendance, and program intensity forms b. Interviews/questionnaires for parents, collaborators, staff	2-3 days
4. <i>Conduct interviews/administer surveys</i>	1-3 days
5. <i>Develop spreadsheets and train staff in data input</i> a. Develop attendance spreadsheet b. Monitor data input	1 - 2 days
6. <i>Aggregate/analyze outcome data and findings from evaluation questions</i> a. Monitor pre-test/post-test data collection b. Tabulate survey/interview/milestone data c. Conduct content and statistical analyses of data	2-4 days
7. <i>Assist in fulfilling state requirements for program quality review</i> a. Develop staff questionnaires/facilitate self study discussions b. Analyze findings/summarize findings	1-3 days
8. <i>Monitor project development/implementation</i> a. Conduct site visits to observe key program activities b. Attend staff meetings	3-5 days
9. <i>Write evaluation reports and recommendations for improvement</i> a. Write executive summary and technical report	4 - 6 days
10. <i>Communicate findings to a variety of audiences</i> a. Develop PowerPoint presentation b. Review findings with staff, collaborators, and administrators	1-2 days
11. <i>Attend state conferences and federal meetings – required to achieve above job responsibilities and activities</i>	1-2 days
The estimated costs provided in the Sample Scope of Work are based on the fewest and greatest number of days. Travel expenses are not included in the following calculations. Range: 20 days @ \$500/day = \$10,000 35 days @ \$500/day = \$17,500	

Note: The sample Scope of Work does not specify a timeline for deliverables. Optimally, a project's final Scope of Work would include a timeline for tasks and deliverables that would reflect state expectations and other Even Start requirements.

Services or activities that are typically not included in a Scope of Work for an independent evaluator include:

- developing outcome instruments (which can be very time- and labor-intensive);
- administering individual or group tests, unless special arrangements have been made;
- conducting personnel evaluations, or using program data for personnel evaluation;
- writing grants; and
- serving as the liaison or primary communication link with the state coordinator.

Establishing costs. As stated, the cost of an evaluation may be pre-specified or it may be a matter of determining the number of evaluator days that are needed to perform the evaluation tasks included in the Scope of Work. When estimating the number of evaluator days, keep the following factors in mind:

- the scope and complexity of the task (e.g., developing an attendance form may be less complex than a parent questionnaire);
- whether the evaluator is providing advice or actually performing the task (e.g., advising on a self study process vs. conducting self study focus groups);
- the number of staff meetings that the evaluator will attend and/or the number of sites where he or she will observe program activities;
- the expense of buying a database or the evaluator's time to develop one, both of which may add considerable cost to the evaluation budget; and
- data collection costs, which will vary depending on the size of the sample and the labor intensity associated with different methodologies.

Projects may consider the following suggestions to reduce the costs of an evaluation:

- Limit the logistics work required of the evaluator. For example, do not ask the evaluator to perform recordkeeping tasks that other staff can readily complete;
- Where appropriate, delegate data collection tasks to project staff and concentrate evaluator expertise on design, analysis, and report writing tasks;
- Use data collection methods that take less labor to implement and analyze. Individual interviews, for example, require more time than paper-and-pencil surveys;
- Prioritize evaluation questions. Select a few for in-depth evaluation each year rather than expecting "full" evaluation coverage of all parts of the program; and
- Use the evaluator for evaluation tasks rather than general program support. Consider carefully any "extras" requests of the evaluator, such as attending staff meetings or accompanying staff to meetings with collaborators.

The Evaluator–Staff Relationship

Orienting evaluators and program staff. Once the project has secured the services of an evaluator, it is wise to schedule an orientation to the Even Start program for all members of the evaluation team. This orientation would generally include meeting members of the staff, observing typical instructional activities and classes, and providing information about key program features and issues. It is helpful to share all types of materials and records, including the program proposal, attendance recording forms, results/recommendations from previous evaluations, and information about evaluation requirements. Information that helps the evaluator understand the program and participants will ensure that he or she does not waste time seeking information that is already available.

The purpose of this orientation is to share information about the program and begin to establish a good working relationship between the evaluators and staff. It is best to keep the orientation informal, for an easy exchange of ideas and suggestions. The next step would be to schedule a more formal meeting with staff and collaborating agencies to begin crafting the evaluation plan. (See Chapter 10.)

Establish regular opportunities to communicate. The strength of the evaluator—staff relationship depends on ongoing and clear communication. That communication will be facilitated by designating a project contact person who will take responsibility for initiating contacts with the evaluator, fielding questions and requests, and conveying information to others.

The intense level of communication that occurs during the design of the evaluation plan needs to be followed with check-in discussions as the evaluator carries out the tasks in the Scope of Work. Regularly scheduled communications such as monthly phone calls or email updates will help alert the evaluator to changes in program operations and let the evaluator share information emerging from his or her activities. It is generally a good idea to have some monthly contact to review accomplishments, discuss upcoming evaluation activities, and adjust the original plan as needed.

Reasonable expectations for a productive working relationship. Working with an evaluation consultant is not that different from working with any other staff member in an organization. The same rules and expectations that govern all collegial relationships apply. From this perspective, it is reasonable to expect that the evaluator will:

- follow through on the evaluation agreement;
- provide timely notification if adjustments to original plans are necessary;
- ask tough questions and challenge assumptions in a respectful manner;
- work in partnership with other staff members;
- respect participants' confidentiality; and
- make suggestions for improving the program and the evaluation.

Monitoring evaluator performance. It is the responsibility of the Even Start project director to monitor evaluator performance and ensure that evaluation tasks are completed according to schedule and at the expected level of quality. The formal agreement for services can be a tool for monitoring performance. If desired, the agreement could specify that final payment be issued after submittal of the final deliverables and a determination that the deliverables have met all state or contract requirements.

As discussed previously, evaluation is collaborative: both project staff and evaluators contribute to its success. On occasions when evaluators do not perform required tasks or provide deliverables on time, project staff members are encouraged to talk with the evaluator to determine the reasons for this lack of performance. For example, an evaluator may not be able to analyze data or write reports because the data the program provided were incomplete, in an unusable format, or late.

Sometimes project staff may disagree with the findings presented in the final evaluation report. If this occurs and cannot be resolved by communicating with the evaluator, project directors are encouraged to provide the project's perspective or explanation as an addendum to the report.

Staff members are encouraged to give regular feedback rather than wait until the end of the agreement to ask questions or express concerns about the evaluator's performance. Regularly scheduled communication can anticipate and resolve many issues before they damage the evaluation.

The success of the evaluator-staff partnership depends on teamwork and cooperation. There may be times, however, when the evaluator and/or the Even Start project decide not to renew the evaluation agreement for another year. One option to anticipate this is to include a clause in the agreement specifying that either party may terminate the agreement with 30 days notice. If the evaluator-staff relationship deteriorates and the project is faced with the decision to terminate or not renew the agreement for the next year, project staff may take solace in knowing that, as the search for a new evaluator begins, this past experience will serve as a source of valuable insight into those evaluator qualifications that best match their program.



Chapter 10

**Planning a
Focused Inquiry**

In addition to analyses of the state required indicators, other project-identified outcomes, and project objectives, local evaluations include a focused inquiry for program improvement. Chapter 10 covers developing a plan to design and conduct the evaluation focused inquiry. Specifically, it presents considerations for:

- choosing the evaluation focus;
- developing project objectives and evaluation questions;
- choosing methods and instruments; and
- developing a timeline.

Earlier chapters examined aspects of participant outcomes and program practices that can guide and inform the evaluation. In practice, a focused inquiry generally addresses one or two aspects at a deeper level, such as:

- exploring reasons for lower or higher than expected indicator results;
- looking more deeply at groups of participants to understand larger patterns;
- finding out more about subgroup services and outcomes;
- gathering multiple perspectives on outcomes;
- supplementing outcomes with more information from schools; and
- gathering longitudinal information.

By focused inquiry we refer to the development of and data collection for key evaluation questions designed for program improvement. The evaluator works with project staff to identify a focus for the inquiry, then takes the lead in shaping the evaluation plan—refining evaluation questions, identifying appropriate methods and instruments, developing a timeline—and carrying out the plan.

In crafting the evaluation plan, the local evaluator draws on his or her evaluation expertise (e.g., informing staff about the trade-offs and consequences of choosing different methods) and facilitation skills (e.g., engaging staff in discussions and decisions about their evaluation). In most instances, the evaluator will have some background knowledge or familiarity with the project. For background on Even Start, the evaluator should consult the Even Start non-regulatory guidance, the Even Start statute, the project's previous evaluation report, and any state Even Start guidance.

Choosing the Evaluation Focus

Sources for identifying the focus include findings from a previous evaluation report (e.g., the average number of home visits is less than intended), changes in participation outcome data (e.g., results were lower than expected and there is no obvious reason why), current issues raised by staff (e.g., levels of reading skills differ by site), and concerns raised by collaborators or stakeholders (e.g., adult participants attend irregularly).

A project's objectives can be another source of focus. Each project has its own objectives, as stated in its approved application or RFP. These objectives are generally based on the Even Start program purposes (Section 1231 [20 U.S.C. 6381] Statement of Purpose) and on the 15 program elements as stated in the federal law (Section 1235 [20 U.S.C. 6381d] Program Elements). A review of progress in meeting the objectives can suggest an area of concentration for the focused inquiry, for example, *Are families staying in long enough to achieve learning outcomes?*

The local evaluator usually leads discussions with project staff to identify an area or areas for the focused inquiry. This discussion can take place at a staff meeting, a special meeting for focusing on evaluation (e.g., a yearly retreat), or with the project's advisory committee. Broad questions such as, *What questions do you have about the program?* can be too vague for staff to answer. More specific questions, such as, *What learning activities worked well for you this past year? What didn't?* are more useful. They may generate discussion about specific areas of concern.

The local evaluator should involve all project partners, if possible, in deciding the evaluation focus and manner of collecting data. The evaluator also may want to involve collaborating agencies in data collection. The more he or she can engage project staff, partnership members, and collaborators in decisions about the evaluation focus and the evaluation process itself, the more likely the findings will be understood and used.

Evaluations of new projects tend to focus on operations and ways to get the project started, such as *strategies for building collaborations*. As the project grows, other concerns arise, such as questions about the effects of program strategies on participant outcomes. Regardless of a program's development stage, the evaluator must keep the focus on improving services to maximize children's and adults' achievements.

If there are several ideas for the evaluation focus, the evaluator and project staff need to decide what is most important and useful to address in the coming year. The evaluator can lead this discussion to choose which questions are essential and which are less so.

Selecting an evaluation focus based on past performance indicator data. Based on past data, project staff may choose to pilot a solution or collect more information about possible solutions. One will be the focus of the evaluation. Examples of choices include:

- Performance indicator data showed that only 14% of parents took the TABE post-test. Further examination showed that the hours offered were insufficient to meet requirements. As a result, project staff used new strategies to increase the number of available hours. The next evaluation can focus on how the new initiatives are working.

Example: Engaging Collaborators

As her first step as an evaluator, Marilyn Muñoz forms an Evaluation Advisory Committee. It includes the project leader, partner agencies and organizations, primary collaborators, staff representatives, and other key players. At the first meeting she introduces evaluation as a way to learn about the project. She leads the discussion by asking questions such as What works well and what doesn't? What are the project's strengths and weaknesses? Where do we need to go back to the drawing board and reconsider? She uses the responses to determine the subject of the focused inquiry and evaluation questions.

Example: First-Year Focus

One evaluator encourages staff to focus the first-year evaluation on collaboration. In addition to child and adult education outcomes, the evaluation addresses getting services running, staff development and training, recruitment, establishing regular communications, etc. Her first-year evaluation also addresses benefits and barriers for collaborators. As a result, collaborators have become more involved with the project during its start-up and have taken an active role in recruiting and serving families.

- Analysis of adult reading achievement showed that adults who do not make progress on the TABE post-test enter Even Start at lower levels than adults who make gains on the TABE. Project staff members try a new strategy: family educators provide adult tutoring or instructional support during home visits to parents who scored low on the TABE pre-test. The next evaluation can focus on the effectiveness of this new strategy.
- Analysis of school age student performance indicators reveals an increase in school attendance but no increase in reading skills, despite the provision of after-school services. Project staff decide to enhance the after-school services, including additional parent-child interactive literacy activity time. The evaluation would concentrate on appropriate activities and their effects on children who are not meeting grade level expectations.

Developing Project Objectives and Evaluation Questions

Identifying objectives. Projects may identify further objectives to help meet performance indicator expectations or other education outcomes. These objectives are in addition to those stated in the approved application or RFP, and may or may not dovetail with the focused inquiry. They also may change every year.

Evaluators can help program staff identify or refine clear and reachable objectives. Some objectives may be connected to education outcomes. For example:

- By October 30, approximately 75% of participating families will demonstrate changes in their parenting and use appropriate literacy strategies to help their children make educational progress.
- By the end of the program year all parents will be surveyed about advocating for their children in school, e.g., their knowledge of parents' rights and responsibilities in the school system.
- By spring, a minimum of 50% of participating adults will show progress toward attaining one or more of their adult education goals.

Other objectives may pertain to project operations, such as

- All staff who provide direct services to families will participate in an average of five hours of general and individual professional development per month.
- Where appropriate, family members, family literacy staff, and other agencies will be involved in setting goals.
- At least 60% of families will remain in the Even Start project for at least one year.

Evaluators can help develop other objectives based on findings from previous evaluations such as recommendations, results from performance indicator data, or staff observations. The evaluator and staff will identify how to measure progress toward each objective. In some cases measuring the objectives will be straightforward; in others, the evaluator may have to ask questions to understand what criteria staff members consider important for achieving their objectives. The evaluator also ensures that there is a means for tracking whether or not the objectives have been met.

Designing evaluation questions for a focused inquiry. Once an evaluation focus has been determined, the next step is to develop relevant evaluation questions. Evaluation questions generally seek information about the value of intended actions or practices. For example, in response to lower than expected findings for children’s reading skills, a project may decide to make strengthening their children’s reading skills part of the parent education curriculum.

General evaluation questions include:

- *How effective is the new focus? What do parents learn? What do they do differently with their children? How do the new skills affect children’s scores on assessments of reading skills? How much do children’s reading scores improve?*
- *Which groups benefit most from the new attention to reading skills? Are parents who are highly engaged in reading to their children more or less likely to apply what they learn?*
- *Under what conditions is the new program element effective?*
- *How satisfied are staff, parents, and collaborators with the new program element?*

In other cases, project staff may decide to collect more information about ongoing practices. Then the evaluator can develop questions that are specific enough to generate usable information from which to develop strategies to strengthen program practices, such as:

- *How does the parenting component affect parents’ interaction with their children’s school?*
- *How does the Even Start project affect parent involvement in the school?*
- *How do oral language development opportunities for Even Start children vary by site?*

Good questions produce answers that can inform decisions. One way to check whether a question will yield useful information is to ask, “*What will the answer or answers tell us about our project? What are possible answers? What will we do if we know. . . ?*” Sometimes one question will be enough, or several questions may be needed.

Another consideration for designing an evaluation question is whether the project administrator and/or staff members can act on an evaluation’s findings. Although a finding may suggest a useful action, the timing may not be right, or authority to act may rest outside the project personnel. For example, findings about the effectiveness of an instructional curriculum may be less useful if staff members can’t influence curriculum decisions. In such cases, the questions could be refined rather than abandoned: the evaluation could address approaches that augment the current curriculum, for example.

Below are suggestions for developing useful evaluation questions⁷:

- Phrase questions so that data can be collected to answer a specific question.
- Allow more than one possible answer to the question; that is, do not use yes or no questions.
- Check that project staff are interested in the questions and care about the answers.
- Confirm that project staff can articulate how they would use the data—ensure that questions are relevant to program operations.

Choosing Methods and Instruments

As the focused inquiry becomes clearer, issues of data collection methods will arise. The evaluator can ask staff members for input on the data collection design and method, e.g., whom to collect data from and what data to collect. Ultimately the local evaluator is responsible for refining the design and identifying the concrete activities. He or she has to make choices about the sample, methods, and instruments.

Sample. The evaluator must identify the appropriate sample. Sample choices are usually limited because the projects are small—the population is often so small the evaluator treats it as the universe, or whole population. Therefore, if the evaluator wants to make inferences or generalizations about the project's population, he or she needs to collect data from all participants, a challenging task. If the evaluator gathers observations from a population subset, he or she should ensure that the subset adequately represents the whole population.

Methods. The evaluator suggests methods for collecting information. Evaluators are encouraged to use relevant comparisons, e.g., similar participants. The Even Start evaluator is not usually in a position to conduct a randomized experimental evaluation study. However, a stronger evaluation will result if the evaluator chooses an appropriate comparison group. Evaluators may use statistical methods to adjust for initial differences between groups.

Comparisons can be made among different groups of participants (e.g., based on attendance hours, entry level scores), among different settings (e.g., observations of the same participants at the home visit and in the classroom), or at different times (e.g., observations of the same participants, in the same settings, at different times).

Evaluators should consider both quantitative and qualitative methods, since the strengths of one can compensate for the weaknesses of the other. Available data sources, such as content analysis of meeting minutes, family portfolios, teachers' observations, etc., can also inform the evaluation. Evaluators should select the most direct methods that provide reliable information, and consider triangulating data from different sources to enhance the dependability of data from small samples.

⁷ From Patton, M.Q., 1997. *Utilization-Focused Evaluation*. 3rd edition. CA: Sage Publications.

Instruments. Because available instruments are usually not specific enough for the project's needs, evaluators sometimes develop protocols to collect data about the local project. These instruments include measures associated with the project's design and implementation as well as participation. The evaluator may develop several instruments to gather information from multiple perspectives, e.g., staff focus group questions, parent surveys, and child interview guides.

Local evaluators may also choose questions from an existing instrument. However, evaluators who adapt existing instruments are encouraged to balance judgments about reliability and validity with the purpose and intended use of the focused inquiry.

Developing a Timeline

The timeline outlines what evaluation activities need to be carried out, by whom, and when. It differs from the Scope of Work and consultant agreement drawn up with the project administrator (see Chapter 9) in containing all the pieces for the focused inquiry, such as identifying the evaluation questions for the inquiry; how, when, and from whom data will be collected; and the timeline for analyzing, interpreting, and reporting the findings. The evaluation timeline also includes analyzing and reporting data the project routinely collects, including participant outcome data, attendance, and other program data. The evaluation timeline is specific enough that staff members know their responsibilities, and broad enough to allow some adjustments in case problems arise. Problems such as delays in getting data from partners, adverse weather conditions, or unexpected staff turnover may arise.

The local evaluator should review the timeline with project staff before finalizing it. During this discussion the evaluator can explain why certain methods were chosen, what trade-offs were made, and probable consequences of the decisions. This discussion may necessitate changes in the evaluation plan timeline. Staff knowledge may influence data collection, e.g., they may have a better sense of parents' and instructors' schedules, or timelines may need to be adjusted in terms of staff commitment.

Reviewing the timeline with staff members allows them to raise questions and concerns about approaches such as a survey versus individual interviews. Throughout these discussions, the evaluator considers alternatives and provides technical advice about the pros and cons of certain decisions. Finally, the evaluator revises the management plan and submits it to the project leader, who makes final decisions about the timeline.

The following page shows an example of an evaluation plan timeline.

Example: Evaluation Plan Timeline

Evaluation Activities	Responsibilities	Timeline
<p><i>Establish evaluation focus</i></p> <ul style="list-style-type: none"> - Meet to discuss evaluation focus 	<p><i>Evaluator, project staff, collaborators</i></p>	<p><i>Early September</i></p>
<p><i>Develop evaluation questions and suggested design and procedures</i></p> <ul style="list-style-type: none"> - meet with staff to discuss plan - revise plan based on discussion 	<p><i>Evaluator & staff</i></p> <p><i>Evaluator</i></p>	<p><i>End of September</i></p>
<p><i>Data Collection Activities</i></p> <ul style="list-style-type: none"> - design and pilot protocols - obtain consent forms if needed - administer pre- and post-tests for state required performance indicators - identify and contact focus group participants - conduct focus groups with program participants - contact teachers for interviews - conduct interviews with teachers - determine how education records will be obtained - collect education records - meet regularly to provide updates 	<p><i>Evaluator</i></p> <p><i>Staff</i></p> <p><i>Staff</i></p> <p><i>Staff & evaluator</i></p> <p><i>Evaluator</i></p> <p><i>Staff</i></p> <p><i>Evaluator</i></p> <p><i>Staff & school collaborators</i></p> <p><i>Staff</i></p> <p><i>Evaluator & staff</i></p>	<p><i>October - February</i></p>
<p><i>Data Analysis</i></p> <ul style="list-style-type: none"> - review and verify data - obtain performance indicator data from project - analyze performance & program objective data 	<p><i>Evaluator</i></p> <p><i>Evaluator & staff</i></p> <p><i>Evaluator</i></p>	<p><i>March – April</i></p> <p><i>May</i></p>
<p><i>Data Reporting</i></p> <ul style="list-style-type: none"> - report & discuss preliminary findings - draft evaluation report - review report - submit final report 	<p><i>Evaluator & staff</i></p> <p><i>Evaluator</i></p> <p><i>Staff</i></p> <p><i>Evaluator</i></p>	<p><i>June</i></p>

The image features three green 3D cubes hanging from a vertical dotted line. The top cube contains the text 'Chapter 11'. The middle cube contains the text 'Reporting and Using Evaluation Findings'. The bottom cube is empty. The entire scene is set against a light yellow background with a green border.

Chapter 11

**Reporting and Using
Evaluation Findings**

Previous chapters of this *Guide* described strategies to help Even Start projects produce useful information for documenting literacy outcomes and improving program practices. The next step is to ensure that the overall evaluation findings are communicated in a manner that encourages Even Start staff and other stakeholders to use them. This chapter covers:

- maximizing the use of evaluation findings;
- the written evaluation report;
- levels of evaluation reporting;
- organization and content of reports;
- enhancing presentations of evaluation findings; and
- reporting data analyses.

Maximizing the Use of Evaluation Findings

Although evaluation is usually just one factor that influences program decisions, the evaluator should ensure that evaluation findings are presented in a way that makes them useful and able to inform decision making. Evaluators should consider the following points in planning for evaluation use:

Audiences. The first step to a useful evaluation is to identify its primary audience. Even Start project staff, partners, key collaborators, and advisory boards responsible for project management and implementation are the primary Even Start audience; they are most likely to act on the evaluation findings (e.g., increase the intensity of a program component, revise curricula, schedule staff development opportunities, improve data collection procedures). Another key information user is the SEA Even Start coordinator, who uses the findings to monitor program progress, make continuation funding decisions, and determine technical assistance and training needs.

Tailored presentation of findings. To maximize their use, evaluators need to tailor their presentation of findings to address what the intended audience wants or needs to know. Different audiences have different information needs and tolerance levels for details. Program staff may be interested in outcome data analyses that explore differences by subgroups or patterns based on program implementation, while state coordinators may simply want to know the percentage of participants who met state performance indicators.

Length of report. Evaluators who present findings without considering the needs or expectations of their audience often present too much information. A key deterrent to using evaluation findings is the mass of material that readers must wade through in order to answer their questions or find information of interest. Because the complete technical report chronicles all evaluation activities and findings, it can be effectively streamlined by including supporting information or additional analyses in appendices. Shorter evaluation progress or interim reports and executive summaries can effectively communicate “need to know” findings to most potential users.

Clear messages. Simplicity and clarity should guide reports of evaluation findings; avoid evaluation jargon and use everyday language. Represent data visually to highlight key findings.

Availability of findings. Timing is everything. Evaluation results that are ready when they are needed are far more likely to be used. Evaluations that address the right issues but arrive “after the fact” may be irrelevant. Evaluators need to balance the thoroughness of their data collection and analysis with the need to deliver accessible findings in a timely manner. “Timely” does not necessarily mean “fast”; rather, it calls on the evaluator to adjust the evaluation schedule so that findings are ready when they will have maximum impact. This may require interim reports or progress memos as needed to inform decisions.

The Written Evaluation Report

Programs should require their independent evaluator to produce a written report of findings and recommendations for improvement. Although there are many options for disseminating evaluation findings, a written report is the most comprehensive format for presenting participant outcomes, answering key evaluation questions, and analyzing data collected from various information sources. Further, a written report has value as a “stand alone” document which serves several important functions:

Accountability. The written report shows evidence of program effectiveness in addressing the state performance indicators and program objectives, and documents compliance with the federal statute. The medium of print formalizes the evaluation and adds credibility and authority to the presentation of evaluation findings.

Historical context. Evaluation reports create a historical record documenting the population served, participant and program accomplishments, and changes in program design over time. This documentation provides an historical perspective and ensures program continuity in cases of staff, evaluator, and/or administrator turnover.

Educating and informing others. The written report promotes understanding. Audiences who are unfamiliar with the Even Start program can use the evaluation report to gain a better understanding of program goals as well as the specifics evaluated e.g., the community context, population profile, program services, and unique features of the Even Start program.

Advocacy and support. By documenting project accomplishments and participant outcomes, the evaluation report is a valuable resource for generating support for Even Start family literacy programs. Evaluation findings can educate policy-makers about the effectiveness of the Even Start model, encourage new collaborations, and enhance program services and operations.

Levels of Evaluation Reporting

The three levels of the written evaluation report are distinguished by the amount and type of information presented, intended audience, and potential use. They are (1) Technical Report, (2) Interim Reports or Progress Memos, and (3) Executive Summary.

The **Technical Report** is the most comprehensive compilation and presentation of information generated by the evaluation. The technical report is often written as an end-of-year report for Even Start programs and is used to meet all of the functions described in the preceding section. Although the technical report is the most informative of all reporting formats, its length and detailed presentation of data tend to limit its audience to the primary users, i.e., Even Start staff and others who implement programs.

Interim Reports or Progress Memos meet the specific information needs of project staff during the course of the project year. Progress reports may also be required by state coordinators and others who monitor program implementation. Interim reports or progress memos are most useful for:

- reporting progress in achieving participant outcomes (e.g., percentage of participants who have achieved project-specified benchmarks related to the state performance indicators at mid-year);
- describing accomplishments and challenges in program operations at a specific time (e.g., number of families served and retention rate, identification of effective strategies to guide staff recruitment efforts); and/or
- presenting findings on a specific area of concern (e.g., results from a parent satisfaction survey on the need to improve program services).

Interim reports or progress memos are typically brief. Depending on the audience and purpose of the report, they can present data formally or informally. Interim reports or progress memos maximize the use of evaluation findings because they have a limited focus and emphasize “need to know” data that are provided while the project is operating and staff or others can take immediate action.

The **Executive Summary** is the short version of the technical report. It communicates essential information about the study’s purpose, design, and key findings without burdening the reader with the complete report. State reporting requirements may specify the length and desired content to be covered. A comprehensive executive summary should contain condensed information from all the major sections of the technical report. It should be written as a stand-alone document that can be included with the technical report or reproduced separately and disseminated as needed.

More focused executive summaries can address an audience’s specific information needs (e.g., state coordinators might request an executive summary of project status with respect to the state performance indicators). Evaluators may need to develop two to three versions of executive summaries, each highlighting the findings of most interest to a given audience.

Organization and Content of Reports

Evaluation findings may be presented in a variety of formats—written reports, oral presentations, and discussions with question/answer interactions. The example below identifies information that may serve the written evaluation report as well as other avenues for reporting Even Start evaluation results. The sample Table of Contents below is followed by more detailed descriptions of information each section of the report could contain. Section headings refer to sections of the sample evaluation report.

Example: Even Start Evaluation Report Table of Contents

I. Program Description

- Program Goals*
- Program Services*
- Partnership Entities and Key Collaborators*
- Participant Profile*
- Staffing*
- Response to Previous Year's Recommendations*

II. Study Design

- State Performance Indicators/Participant Outcomes*
- Project Objectives*
- Evaluation Focus and Key Evaluation Questions*
- Evaluation Design: Data Sources, Sample, and Methods*

III. Evaluation Findings

- Participant Outcomes: State Performance Indicators*
- Other Participant Outcomes*
- Program Participation: Attendance and Retention*
- Findings Related to Key Evaluation Questions*
- Interpretation/Discussion of Findings*

IV. Conclusion and Recommendations

- Summary of Findings*
- Progress based on Previous Year's Recommendations*
- Recommendations for Improvements*

Appendices

- Instruments*

Part I: Program Description. Information about a program's goals, services, and characteristics of the participants served is essential for ensuring that all readers understand the context of the program being evaluated. This information also helps the reader interpret the evaluation findings. We recommend that all presentations of evaluation findings begin with a program description (one to two pages in the technical report) or highlights of program features (Executive Summary or PowerPoint presentation). The table on the next page presents key content the program description could include.

Example: Program Description

Program Goals	<i>The CV Even Start Family Literacy program completed its fourth year of operation as a cooperative effort between the Library and its co-applicant, the School District. The Even Start program goal is to improve the educational opportunities of families by 1) helping parents become full partners in the education of their children, 2) assisting children in reaching their full potential as learners, and 3) providing workforce and literacy training for parents.</i>
Program Design	<i>The program design incorporates aspects of both a home-based program and the Kenan Family Trust center-based model. As part of the center-based program design, parents attend adult education classes while their children attend a Head Start preschool program or an enriched childcare program, which is available for children ages birth to 5. Parents also participate in parent education class/support groups and parent-child interactive literacy activities. Families receive instructional support services twice a month during home visits.</i>
Program Services	<i>An intensive year-round program, CV Even Start offers 600 hours of ESL instruction and 300 hours of GED preparation. The Head Start preschool program offers 800 hours of early childhood education. Families may also participate in 150 hours of parenting-related classes or activities over the course of the project year.</i>
Partners & Collaborators	<i>The Library serves as the fiscal agent and provides a variety of services for staff and participants, including free book distribution, individualized tutoring, and small group literacy instruction. The district supports the program by providing facilities, staff training, and early childhood education programs. The co-applicants established a collaborative agreement with the High School District Adult School for adult education services and childcare.</i>
Participant Profile	<i>The project serves low-income, Hispanic families living in the attendance areas of three elementary school schools: VS, LV, and HS. During the 2002-03 school year, the project served a total of 32 families (32 adults and 48 children). Of the families enrolled in the Even Start program, three (7%) received TANF and four adults (10%) were employed. Nearly two-thirds of the families (61%) reported living on an annual income of \$19,999 or less. None of the adults possessed a high school diploma.</i>
Staffing	<i>A full-time family literacy director coordinates the program. The program is staffed by three instructors responsible for developing and delivering the home-based instructional programs. Instructional staff (provided through in-kind contributions) includes ESL/GED instructors (Adult School) and preschool teachers (Head Start). All instructional staff members possess the minimum required qualifications for their positions.</i>
Unique Features	<i>The program is designed to serve the literacy and employment needs of families participating in the Welfare-to-Work program and offers individualized services to accommodate the schedules of working adults.</i>

Part II: Study Design. This section describes the key features of the evaluation plan. It should include information on the study's goals, focus, assessments, and data collection design. The description of the study's design helps the reader determine whether the evaluation was methodologically sound and likely to produce credible findings. This section of the technical report may be brief (3-4 pages) but should include sufficient information to communicate:

- Purpose of the evaluation study (e.g., evaluation goals, target audience and expectations for how the data are to be used, reference to evaluator/agency conducting the study);
- State performance indicators, project objectives, and other participant outcomes (e.g., lists, summaries, or references to participant outcomes that are the focus of the study);
- Focus area(s) of inquiry (e.g., focus area(s) and project-identified evaluation questions to be answered as part of the study); and
- Data collection plan (e.g., assessments used to measure outcomes and gather information related to evaluation questions, sample, methods for data collection).

Part III: Evaluation Findings. The findings section of the evaluation report presents analyses of the data and describes findings on family participation patterns, participant outcomes, and evaluation questions. (Suggestions for analyzing and reporting specific types of data are provided in Chapters 3-7 of this *Guide*.) The findings section is also the appropriate place in the report for the evaluator's comments on the quality of the data and any background information that may help the reader understand issues that may have affected data accuracy or completeness (e.g., due to a late start-up date, the project is reporting pre-/post-test data based on five months of program implementation). See below in this chapter for a discussion of presentation of findings.

Part IV: Conclusion and Recommendations. The final report should close with a conclusion and recommendations for improvement. The conclusion concisely summarizes key findings, answers the key evaluation questions, and draws conclusions about the program's effectiveness in achieving outcomes and state performance indicators. As stated above, evaluators are encouraged to be brief and to present their summary statements clearly and simply. Presentations of evaluation findings should end with a conclusion (one to two pages in the technical report) or summary statements (Executive Summary or PowerPoint presentation). The example on page 106 shows key content the conclusion could present.

Recommendations for Improvement. All Even Start evaluation reports should include recommendations for improvement. They are the formal link between evaluation findings and their use for program improvement by staff and others. Methods for developing recommendations are varied and reflect the extent to which staff and other stakeholder have been involved in the

Example: Recommendation for Improvement

Comparisons of students' mean pre- and post-test CASAS reading scores show little evidence of student growth in reading. The ABE Performance Indicator states that 50% of students will make a three-point post-test gain on the CASAS reading sub-test. Only one student (6%) achieved the performance indicator.

In response to last year's recommendations for improvement, the program increased the number of instructional hours offered in adult education. Currently the class offers 225 hours of instruction; students attended an average of 135 hours (60% attendance rate).

It is recommended that the program increase the intensity of the adult education program offerings to a minimum of 60 hours a month and develop incentives to encourage students to attend more consistently. The adult education teacher may also wish to review the adequacy of the current reading curriculum for adults with learning disabilities and examine lesson plans to ensure that sufficient time is allocated to reading activities.



Example: Conclusion

Project/ Family Summary	<i>The Even Start program has successfully operationalized the four components of family literacy in a home- and center-based program and served a total of 50 families (50 adults and 80 children) during the 2001-2002 school year. These families represent an extremely high need population: average family income is less than \$11,000; fewer than half the adults have completed nine years of schooling; and one-third of the children between birth and age 3 have been diagnosed with severe developmental disabilities.</i>
Attendance/ Retention Summary	<i>The family retention rate for the school year is 75%, including seven adults who achieved their goals and graduated from the program. This highly intensive program provides almost 650 hours of ESL instruction and over 300 hours of parenting-related classes or activities. Most families met the program expectation that they would attend a minimum of 70% or more of the classes offered within each component.</i>
Performance Indicator Status	<i>Findings from the fourth-year evaluation study showed that, with the exception of the indicator related to GED attainment, participants met and exceeded the state performance indicators for adults and children. Of the 10 adults enrolled in GED preparation classes, two (20%) successfully completed their studies. (Indicator specified a 50% completion rate).</i>
Evaluation Question (Focused Inquiry)	<i>The key evaluation question examined the effect of home visits on family participation and retention rates. The study findings indicated that families who received two home visits a month participated at a higher rate and remained in the program longer than families who received only monthly or quarterly visits. Families who received more frequent home visits attended an average of 80% of their class offerings (compared to 65% for families visited monthly). All “two-visit/month” families are still enrolled in the program.</i>

evaluation study. Recommendations may be developed collaboratively by staff and the evaluator, or they may be developed by the evaluator and presented to staff for discussion and modification. Evaluators who involve staff and other stakeholders in developing recommendations ensure maximum ownership of the findings and agreement with any suggested corrective action. Below are qualities that evaluators and project staff should consider in developing their own recommendations for improvement:

- The problem should be well defined and supported by data in the findings section of the report.
- The recommendation should be action-oriented and offer a variety of potential solution strategies.
- Multiyear evaluations should include a progress report on recommendations proposed in previous studies.

Enhancing Presentations of Evaluation Findings

We encourage evaluators to enhance their presentations of findings with visual representations of both quantitative and qualitative data. Tables and figures are effective ways to communicate evaluation findings because pictures of patterns, trends, and relationships help the reader to assimilate information more easily. Visual or graphic forms are also efficient: they can present a considerable amount of data in a limited space. The evaluator should ensure that the visual representations he or she chooses are the most accurate, complete, and effective presentation of the findings. Readers of evaluation reports are cautioned that visual representations, although often pretty and compelling because of their

apparent simplicity, must be used and interpreted with care. The following are general tips for using tables and figures to represent data visually.

- Consider the nature of the findings and select the most effective type of visual (e.g., tables, charts, or graphs) for presenting data.
- Be selective. Don't overload images with information.
- Prepare the reader by describing the purpose of the tables, charts, or graphs.
- Title the visual clearly and label all rows, columns, sections, etc.
- Indicate the number of participants that the data represent.
- Indicate what the scores represent (e.g., standard scores, percentiles) and include a written interpretation of the data.
- Recognize that visuals alone cannot tell the whole story.
- Recognize the limitations of visuals, e.g., over-simplification of data, selective presentation of findings, which can lead to inaccurate conclusions.

Using tables to present data. Tables present numbers or text in rows and columns. Although tables present large quantities of data concisely, they are less valuable for showing trends or illustrating data patterns. Tables are most useful when it is critical that the values are displayed accurately (e.g., differences in pre- and post-test mean scores). For example, in Even Start evaluations, tables can effectively display participant profile data, attendance and retention findings, and results from pre- and post-test assessments. (See sample tables below, in this chapter). To ensure that tables convey all the information necessary for interpreting the data, evaluators are encouraged to include:

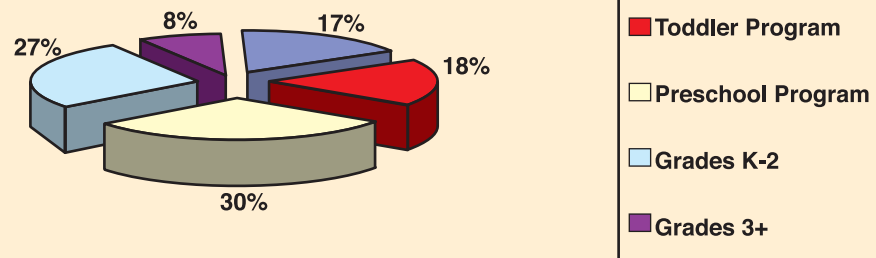
- An introductory paragraph describing the table's purpose and the content covered (e.g., the assessment and interpretation of scores);
- A title that includes all pertinent information (e.g., type of data, comparisons of different groups, time period, assessment used, target group represented, number of participants);
- Labels for each column and row; and
- A concluding paragraph that summarizes the key findings and notes the statistical and practical significance of the findings, where appropriate.

Figures—charts, graphs, or other illustrations—are useful for presenting quantitative or qualitative data visually. Figures are effective because they provide “pictures of the data,” making statistical analyses of complex data quickly and easily understood by all audiences. There are three primary forms of figures that evaluators might include in their evaluation reports.

Pie Charts show 100% of a variable divided into different parts or slices (of a pie). Pie charts can be used to present quantitative or qualitative data and are most useful for showing the relationship of the parts to the whole. Pie charts can effectively present information about participants and program services.

Pie Chart

Enrollment Data: Child Participants in Early Childhood Education Classes



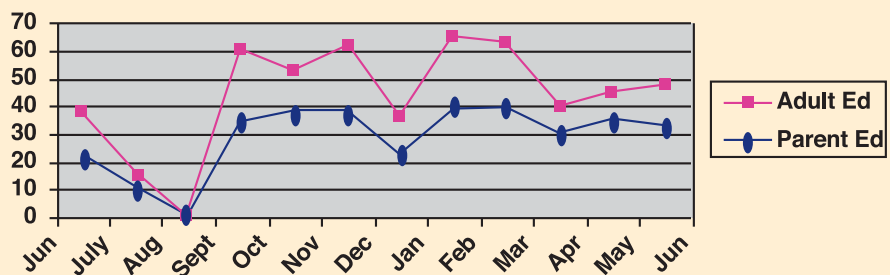
Total Number of Children Served = 78

Bar Graphs use the symbol of a bar to display data and show status at different times, such as participants' reading test scores at pre- and post-testing, or different variables for the same time period (reading, writing, math scores at post-testing). Bar graphs are particularly useful for illustrating achievement of performance indicators and displaying analyses that explore comparisons, patterns, or trends in the data. (See example of a bar graph on page 110.)

Line Graphs convey information plotted on a graph. Line graphs can show results from two or more variables across time and illustrate trends or changes in the data (e.g., participant attendance by program component).

Line Graph

Average Number of Adult Education and Parent Education Hours Attended by Month



Reporting Data Analyses

Reporting participant outcome data. A key purpose of an Even Start evaluation is to show evidence of participant progress in achieving state and other project-identified outcomes. At a minimum, the evaluation report should present analyses of participant outcomes on state performance indicators. The report should offer statistical analyses of assessment results, reference each performance indicator, and include written comments to help the reader interpret the findings.

Example: Participant Outcome Results

ABE Performance Indicator: Fifty percent of adult learners enrolled in Adult Basic Education classes who achieve a pre-test scale score of 211 or higher on the CASAS Reading and Math Tests will demonstrate a three-point post-test gain in each subject area after a minimum of 100 hours of instruction.

The Comprehensive Adult Student Assessment System (CASAS) reading and math subtests are administered on a pre- and post-test basis to adults who completed a minimum of 100 hours of ABE instruction over the course of the project year. All enrolled students were administered the CASAS pre-tests (n=30); 70% (n=21) completed the reading post-test and 63% (n=19) completed the math. Table 1 shows participants' mean pre- and post-test scores and the percent that achieved the ABE Performance Indicator.

Table 1
CASAS: Analyses of Matched Pretest/Post-test Scores

Subject	Number	Mean Scaled Score		Mean Difference	Percent Achieving ABE Performance Indicator
		Pretest	Post-test		
Reading	21	237.38	243.85	+6.47*	18 (86%)
Math	19	227.52	231.68	+4.16	12 (63%)

**Difference is statistically significant (p < .01)*

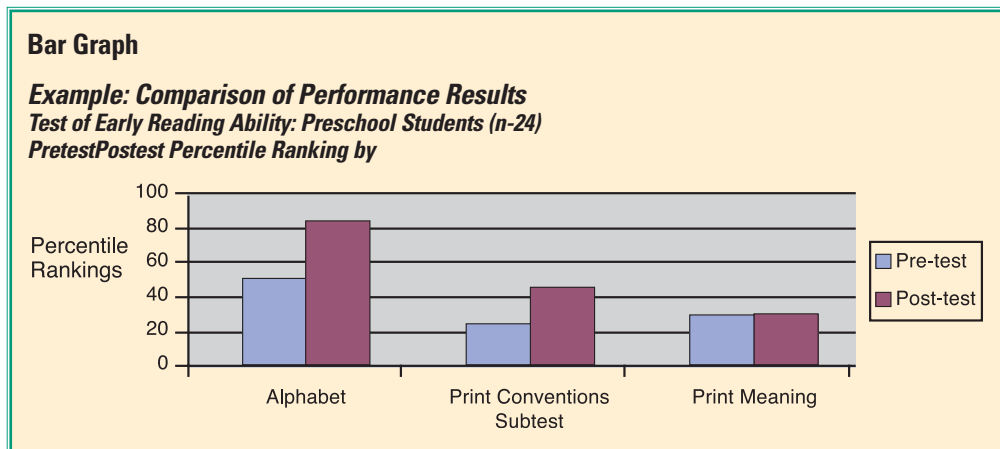
Summary: *The project met and exceeded the ABE Performance Indicator (86% achieved the reading indicator and 62% achieved the math performance indicator). At post-testing, 14 students scored at the highest level of the CASAS reading test (Advanced Adult Secondary) which suggests that they “can comprehend college textbooks and apprenticeship manuals.” The increase in scaled scores from pre- to post-testing is statistically significant, meaning that the difference is not attributable to chance.*

Interpretation of Table: Row 1 – Reading. *Twenty-one participants were pre-tested on the CASAS reading test and then post-tested after a minimum of 100 instructional hours. The pre-test mean (average) scaled score for these participants was 237.38; the mean post-test score was 243.85. The mean post-test score is 6.47 points higher than the mean pre-test score and the difference between scores is statistically significant. Of the 21 participants, 18 or 86%, achieved the performance indicator of a 3-point gain.*

Reporting supplemental analyses of participant outcomes. In addition to presenting data specific to the performance indicators and other project-identified outcomes, evaluators are encouraged to analyze the data further to explore patterns or trends. Suggestions for supplemental analyses include:

- Relating outcomes to each other (e.g., analyzing differences in achievement on literacy/language subtests or other areas of development);
- Relating outcomes to program participation (e.g., analyzing achievement patterns based on contact hours or length of enrollment in program);
- Comparing outcomes to previous year's outcomes (e.g., documenting differences in achievement patterns over time);
- Looking more deeply at individual cases to understand larger patterns (e.g., reviewing family case histories to find trends or patterns influencing achievement or lack of progress); and

- Conducting comparison studies of sub-groups (e.g., participants receiving services from different providers).



Reporting program participation data: attendance and retention. Family participation in program services is a critical variable in achieving participant outcomes. As such, it is helpful to include findings describing participant attendance patterns and retention rates in the evaluation report. This section could include analyses of adult and child attendance data for the key program services offered over the project year, the percent of participants who met the project's attendance expectations, and annual family retention rates.

Example: Participation Data
Adult Attendance by Component: Median Number of Hours, Range, and Attendance Rate

Program Component	Number of Participants	Median Number of Hours	Range of Hours	Average Attendance Rate*
Adult Education: ESL	33	462	124-717	88%
Parent Education	33	176	77-225	75%

* Project Attendance Policy: Participants attend a minimum of 75% of their classes in each component.

Example: Participation Data
Number of Families Served: 44
Number of Families Dropped from Program: 16

Reasons why 16 families exited the program prior to attaining their goals.

- | | |
|----------------------------------|------------------------------|
| 1) Moved out of the area: | N=6 (6 participants or 37%)* |
| 2) Illness/Mental Health Issues: | N=5 (31%)* |
| 3) Reasons Unknown | N=3 (19%) |
| 4) Poor Attendance | N=2 (13%) |

** During the 2002-03 project year, the program recruited families involved in the court system. These families were not sufficiently stable to participate fully in a family literacy program.*

Reporting findings from a focused inquiry. The findings section of the report should conclude with a presentation and discussion of data that address the key evaluation questions in the focused inquiry for that project year. This section may present more in-depth analyses relating outcome findings to program implementation and/or findings from additional data collection efforts (e.g., survey or interview findings on program implementation concerns, participant perceptions of program services, achievement based on established milestones).

Example: Focused Inquiry Results

Key Evaluation Question: *How have parents changed their parenting practices at home to support the development of their children's literacy skills?*

In individual interviews with the program evaluator, a randomly selected sample of parents (n=15) offered their perspectives on changes in their parenting behaviors. In response to an open-ended question asking parents to describe if they did anything differently at home after having participated in the program, parents reported that they a) read on a daily basis (n=13), b) visited the library more frequently (n=11); c) engaged in interactive reading behaviors (n=10); and d) helped children with homework (n=8). Selected self-reported changes in parenting behaviors are provided below.

"Now I go to the library and I try to read to my children every night. Before, I didn't read at all."

"I don't watch soap operas anymore. Instead of spending one hour watching TV every night, I spend the time playing word games with my girls. We play together and then read books."

"I read more to my daughter. I ask her questions about the books we read. I never did that before the program. Also, I never helped my older child. Now I do homework with him."





Chapter 12

**The State Coordinator's Role
in Local Evaluations**

This chapter identifies state leadership actions that can influence the quality of evaluations at the local level. It describes five goals for state coordinators:

- establishing and communicating state requirements and policies;
- providing information, training, and support to evaluators;
- training project staff in evaluation concepts;
- reviewing and offering feedback on evaluation plans and reports; and
- using the local evaluation results.

While the other chapters in this *Guide* are targeted to local evaluators and project directors, state-level leaders also create conditions for high quality and useful local evaluations. This chapter encourages state coordinators to review their options for improving local evaluation practice.

The ideas in the *Guide* are relevant to all states—even those with only a few Even Start projects. Some of the ideas may require that the Even Start state coordinator use state-level administrative or technical assistance funds to add the expertise of a state evaluator, such as a staff member skilled in evaluation or an outside contractor, to support local evaluation activities and conduct statewide activities such as summarizing results from local programs. State coordinators may suggest to local projects a single evaluator to conduct many local evaluations. (It is important to remember, however, that local evaluations need to have specific local relevance, including customized information for program improvement.)

A state's ability to implement some strategies may be influenced by the size of the state allocation. Some will need creative solutions such as joining with related state agencies or Even Start coordinators from other states to accomplish improvements. All states should consider acting on these five goals to support high quality local evaluations:

1. Establish and communicate state requirements and policies for local evaluations;
2. Provide information, training, and support to evaluators,
3. Provide training in evaluation concepts to project staff;
4. Review and provide feedback on evaluation plans and reports; and
5. Use the results of local evaluations.

Each goal is discussed below with examples from several states. There are many ways to address each goal, of course, and states will tailor actions to state circumstances.

Establishing and Communicating State Requirements and Policies

Strengthening local evaluation practice begins by clarifying expectations and communicating a vision of local evaluation to program directors and evaluators. Local evaluators and the project staff must know what is expected.

States may elect to formalize evaluation requirements through policies that govern how projects select evaluators and what evaluators are expected to do. For example, a state may govern evaluator qualifications, even providing a list of approved evaluators from which projects select. Or a state may define expected evaluation tasks and the level of compensation. It has become common for states to require evaluators to summarize and report data relevant to state performance indicators. Some states operationalize requirements in the form of report templates that guide data collection and reporting. State coordinators should ensure that projects spend enough resources on evaluation to do an effective job of collecting data for evaluation questions and analyzing outcomes to guide program improvements.

To clarify expectations, state coordinators can:

- define the expected roles of local evaluators in written policy, specifying the intended functions of local evaluation and the minimum expectations for all local evaluations;
- provide information about the range of costs typically required to meet the minimum expectations;
- provide general information about expected evaluator qualifications while not being overly prescriptive;
- allow evaluators flexibility to include evidence for questions that are important to the local context; and
- provide guidance for what information should be reported annually.

The focus of local evaluation should not be narrowed to collecting and reporting only performance indicator data. However, states should define the evaluator's expected role with performance indicators. They should state expectations for quality control associated with data collection for the performance indicators; analyses that go beyond simply summarizing and reporting information; meaningful comparisons of results of performance indicators with other data; and/or identifying recommendations emerging from the indicator results. Monitoring the degree to which local projects comply with legal requirements is not the evaluator's role.

Evaluation vs. Monitoring
Supporting local evaluation is different from monitoring local projects. State coordinators monitor the degree to which local projects comply with legal requirements, report on the state's performance indicators, and fulfill their approved objectives. In such monitoring, the state coordinator may use some of the same tools that evaluators use, e.g., a statewide database to track participant attendance, but the purpose of monitoring differs from that of local evaluation, which is to collect evidence to inform improvements in the local program.



Example: Common Framework of Expectations

One state coordinator worked with a small group of experienced evaluators to develop a framework for local evaluations. The framework specified the evaluator's role in overseeing the summary and reporting of performance indicator data, designing and conducting an annual focused inquiry based on one or more research questions, and making recommendations for program improvement. The coordinator presented the framework to the state's Committee of Practitioners to be approved as Even Start policy, and the framework accompanied the request for new grant proposals and documentation required for continued funding.

Example: Targeted Training Opportunities

One state allots a special day for evaluators in each statewide training conference. Agendas are planned by an evaluation consultant working with local evaluators. Local evaluations are presented as cases, along with sessions on issues of interest, e.g., statewide results, selecting early childhood progress measures, ways to summarize data, importing data into spreadsheets, etc. As a result, evaluators have become colleagues and tap each other's expertise throughout the year.

Providing Information, Training, and Support to Evaluators

Evaluators usually want more information from states about Even Start operations, evaluation expectations, performance indicators, best program practices, and evaluation design options. They appreciate the opportunity to meet with fellow evaluators and learn how others approach data collection. Some states have an annual cycle of evaluator training.

State coordinators should consider the value of:

- providing at least one annual training opportunity for local evaluators;
- providing annual training in the state's required performance indicators and reporting system;
- encouraging evaluators to attend appropriate portions of statewide Even Start meetings (and encouraging projects to include time to do so in their agreements with evaluators);
- including evaluators on project listservs so they can keep up with changes in Even Start;
- offering orientations or mentorships for new evaluators to provide information such as that contained in this *Guide*; and
- presenting evaluators with information about statewide results for comparative purposes.

States cannot provide complete training in basic evaluation methods, but rather demonstrate the application of evaluation methods to Even Start circumstances. State coordinators should engage someone with evaluation expertise to help design and provide training. The person selected should have a neutral perspective and the respect of the state's professional evaluators.

Training Project Staff in Evaluation Concepts

Even Start project leaders may have no prior experience with hiring independent evaluators, developing an evaluation plan, or using the results of evaluations. Lack of knowledge may put them in an awkward position during hiring and may limit their abilities to ask for—and actually receive—what they need from an evaluator. State coordinators are best positioned to provide “consumer” information, but they need to do so before projects have engaged evaluators and evaluators have begun collecting data.

State coordinators should consider providing:

- information about evaluator qualifications in the request for proposals;
- information about evaluation expectations in orientations for new projects;
- this *Guide* or similar materials to all new projects;
- cost guidelines for evaluations;
- any applicable state procurement policies or hiring guidelines;
- sample evaluation agreements with local projects;
- model evaluation reports as well as criteria for determining the quality of evaluations;
- and
- clarifications about the differences between advocacy and evaluation for continuous improvement, and reinforcements of those distinctions at every opportunity.

It is as important to inform local projects about using evaluation findings and their responsibilities in seeking useful information as it is to address the technical aspects of hiring an evaluator. Hosting sessions that evaluators and project staff members attend together can help raise issues associated with working together, although it is also a good idea to allow each group to meet separately. State coordinators should work with projects to change evaluators when the evaluator is not generating useful information and support has been provided to him or her.

Reviewing and Offering Feedback on Evaluation Plans and Reports

State coordinators who read and react to evaluation reports in a timely manner demonstrate the importance of producing quality evaluation reports. The state coordinator's attitude about the value of evaluation will come through in the attention paid to evaluation results—an attitude that can influence the value that local projects place on evaluation. By paying attention to the findings reported by evaluators, state coordinators communicate how much they value the evaluator's independent perspective.

A state coordinator's reviews can have different purposes: looking for evidence of best practices to be shared with other projects; identifying project needs that can be addressed with additional resources or supports; identifying issues that require more information; and/ or identifying strengths and weaknesses of the evaluation approach. Feedback in any of those areas can reinforce the importance of solid local evaluations.

Of course, it may not be possible for the state coordinator to review all evaluation reports personally, but systems can be set up through which all written reports receive some attention and feedback.

Example: Getting Off to a Good Start

In one state, the request for new proposals includes the state's policy on local evaluation and a one-page description of what is expected of the local evaluation during the project's first year of operation. Projects can go to a website for information about Even Start, including examples of local evaluation reports, the state's performance indicator requirements, and last year's statewide indicator results. Numerous directories, including a directory of evaluators, are available online.

Example: Team Review of Reports

Local evaluators work with a team from the state agency to review local evaluation reports. Each report is reviewed by a team of two who follow a protocol for providing written feedback about the quality of the evaluation plan and execution. The teams can review and write feedback on three or four reports each day. One goal is to identify strong models that can be shared throughout the state.



Example: Troubleshooting with All Parties

Based on this year's review of performance indicator data, the state coordinator is concerned about three projects from the standpoint of limited participation by families and disappointing results for families that are reported. The Stoneway Even Start project is one of the three.

The state coordinator has just received Stoneway's annual evaluation report which indicates, among other issues, that none of last year's recommendations were implemented due to staff turnover. On the bright side, the report also describes progress that adults are making on learning skills using a pilot curriculum with its own assessments.

The state coordinator calls Stoneway's director and asks for a meeting that includes the project evaluator, leaders from partner agencies, and lead staff to discuss expectations for the coming year.



The state coordinator might share the review of reports with colleagues in the state agency who work in cooperating programs such as adult education or with independent consultants. Peer reviewers—other local evaluators or project directors—can be ideal reviewers. Peer reviewers should use a structured process and ensure that confidentiality is respected and a spirit of improvement characterizes the review comments.

State coordinators should:

- establish the expectation of receiving an evaluation report from each project at least annually;
- find a way to give some feedback to project leaders and evaluators about all evaluation reports each year;
- consider that the primary function of evaluation reports is an analysis of information that the evaluator has collected for improving the local project;
- periodically share the strongest examples emerging from reviews;
- provide general feedback about the patterns observed in the reviews; and
- develop a strategy for working with evaluators whose work falls short of expectations.

Using the Local Evaluation Results

The state coordinator's use of local evaluation findings conveys the most powerful message about the utility of local evaluations. In some states, coordinators rely on local evaluation reports as the source of data about performance indicator results—automatically elevating the importance of the evaluator's role in contributing to Even Start's effectiveness. In others, state coordinators have found ways to weave the usefulness of local evaluations into a range of operations.

State coordinators should consider:

- including the evaluator in their on-site program visits and incorporating questions from the local evaluation in discussions with project staff about progress;
- acknowledging to the local project director and the evaluator the receipt and review of each year's report;
- including in their policies on local evaluation the expectation that after a project's initial year of operation, each evaluation report will include information about responses to recommendations made in previous years' evaluation reports;
- using the local evaluation as a source of information when determining which programs most need technical assistance;

- sponsoring sessions at state conferences about the results of particular evaluations that are relevant to many programs; and
- incorporating information from local evaluations in statewide reports.

These recommendations may require a change in culture in some states toward a more public and explicit use of local evaluation results to strengthen family literacy programs.

State coordinators who have already taken action in each goal area should share with other states their ideas for improvement strategies, including ways to fund activities. The ideas above may inspire ways to build on existing foundations.

For states that have few strategies in place, implementing ideas from several goal areas should speed the rate of visible improvements in evaluations and programs. Other state coordinators are an important resource and can augment the ideas presented in this chapter.



Glossary

This glossary is intended for the purpose of clarifying items in the *Guide*.

Adult education: One of the four core instructional components that local Even Start projects must provide as part of “family literacy services.” The adult education component provides high-quality, intensive instructional programs in adult literacy or adult basic education, including English language literacy as necessary, that leads to economic self-sufficiency. (Sections 1231(1), 1235(4) and 9101(20), ESEA; 20 U.S.C. 6381, 6381d(4) and 7801(20).) Section 1235(10) of the Even Start statute (20 U.S.C. 6381d(10)) requires adult education services to be based on scientifically based reading research to the extent that research is available.

Attendance patterns: Information about how and when participants attend instructional components, determined by analyzing participant attendance data. Attendance patterns help determine whether participants are attending enough class hours for a long enough time to achieve their program goals.

Attendance rate: A percentage based on the number of hours a participant attended program services out of the total number of instructional hours he or she could attend. Attendance rates help determine whether participants are attending program services consistently.

Bar graph: A graph that uses a bar to display data for variables, including groups over time. The height of the bar shows how many things or people are in a group.

Benchmark: A standard or point of reference (often a milestone related to an outcome).

Collaborator or collaborating agency: As used informally in the Even Start context, a collaborator or collaborating agency or organization generally means an organization, agency, or entity that may provide program services but is not a formal partner in the “eligible entity” partnership that receives the sub-grant from the SEA.

Continuous improvement: The use of evaluation data to make sure that program instruction and services keep improving over time.

Core instructional components: All Even Start programs are required to provide instruction in 1) adult literacy; parent literacy training that leads to economic self-sufficiency (*adult literacy or English language instruction*); 2) age-appropriate education to prepare children for success in school and life experiences (*early childhood education*); 3) training for parents to be the primary teacher for their children and full partners in the education of their children (*parenting education*) and 4) interactive parent-child literacy activities.

Criterion-referenced tests: Standards-referenced tests or proficiency tests that measure how well a student performed against a certain criterion or standard.

Curriculum: The sequence of instructional content and skills designed to ensure that students understand and are able to achieve proficiency on standards.

Disaggregation: Separating data into groups by variables (e.g., attendance rate, length of time in program, age, pre-test status) in order to see patterns.

Early childhood education services: One of the four core instructional components that local Even Start projects must provide as part of “family literacy services.” The early childhood education instructional component provides high-quality, intensive, and age-appropriate early childhood educational services to prepare children for success in school and life experiences (Sections 1235(4) and 9101(20), ESEA; 20 U.S.C. 6381d(4) and 7801(20)). Section 1235(10) and (12) of the Even Start statute (20 U.S.C. 6381d(10) and (12)) require early childhood education services to be based on scientifically based reading research to the extent that research is available and to include reading-readiness activities for preschool children based on scientifically based reading research to the extent available, to ensure that children enter school ready to learn to read.

Evaluation question: A clear statement in the form of a question to elicit information that an evaluator and program staff wish to study.

Executive summary: A brief, condensed version of a technical report. Drawing on all major sections of the report, an executive summary may be a stand-alone document, either included in front of the technical report or reproduced separately and disseminated as needed.

Expected outcomes: The effects on participants that can be anticipated as a result of full participation in a program that is implemented as intended.

Fidelity: Delivering actual services according to their original design.

Focus group: A group of people who have shared an experience (for example, parents of school age children) and are asked to reflect on that experience through a semi-structured discussion.

Focused inquiry: A systematic exploration of a program interest or concern. The evaluator and program staff develop one or more guiding questions for the inquiry; the evaluator collects and analyzes data.

GPRA indicators: By requirement of the Government Performance Results Act (GPRA), each federal executive agency has established sets of statements of measurable outcomes.

Grade-equivalent or developmental age score: The school grade level or chronological age for which a given score is the estimated median or mean.

Independent evaluation: A program evaluation which uses an independent evaluator (also referred to as an outside evaluator or an external evaluator) to provide an objective and unbiased report of findings for program improvement. The statute requires each local Even Start program to have an independent evaluation.

Independent evaluator: A consultant to the Even Start program, the independent evaluator generally should not be an employee of a partner responsible for administering the Even Start project. An exception would be an employee of a large school district or institution of higher education who may serve as the project evaluator provided that he or she does not work in the department that administers the Even Start program or provides program services.

Instructional approach: The combination of materials and teaching methods or pedagogy that are used to deliver the curriculum.

Instructional opportunity: The number of instructional hours offered during the project year (see program intensity).

Instrument: A measure or protocol designed to capture information from respondents in a systematic way. An instrument can include tests of knowledge or ability, surveys of perceptions, rating scales to assess satisfaction, and so forth.

Interactive literacy activities (ILA): Interactive literacy activities between parents and children (formerly called parent and child activities [PACT]). One of the four core instructional components that local Even Start projects must provide as part of “family literacy services.” The interactive literacy activity (ILA) instructional component provides high-quality, intensive, interactive literacy activities between parents and their children. (Section 1235(4) and 9101(20), ESEA; 20 U.S.C. 6381d(4) and 7801(20).) Section 1235(10) of the Even Start statute (20 U.S.S. 6381d(1) requires instructional components to be based on scientifically based reading research, to the extent that research is available.

Interim report: A periodic report to meet a project’s real-time information needs before a final report is complete. The interim report may 1) report progress toward achievement of participant outcomes, 2) describe accomplishments and challenges in program operations at a specific point, and 3) present findings on a specific area of concern.

Learning milestones: Key achievements that mark progress toward longer term educational outcomes.

Likert scale: A rating scale measuring the strength of agreement with a statement, usually four or five points on a scale represented by a number.

Line graph: A figure that conveys information that is plotted on a graph. A line graph can show results from two or more variables across time.

Mean hours of attendance: The average number of hours of attendance. The mean is obtained by dividing the sum of participants' attendance hours by the total number of participants.

Median attendance rate: The midpoint in a series of numbers of hours that participants attended. Evaluators often analyze median rather than mean hours because the median better reflects the wide **range** of hours attended.

N: The number of people or things in a group that were studied, expressed as N= (e.g., N=15)

Norm-referenced test: Any test for which the scores are compared to the scores of a "norming group" (a representative cross-section of subjects who have taken the same test). Norm-referenced tests show how the test-taker's score compares to the distribution of scores. (For a test-taker at the 50th percentile, it means 49 out of a hundred test-takers would have scored lower than the test-taker.)

Outcome measure: The measure of expected change—usually the difference in scores before and after an intervention or the attainment of particular criteria.

Parenting education: One of the four core instructional components that local Even Start projects must provide as part of "family literacy services." The parenting education instructional component provides high-quality and intensive instructional programs for parents about how to be the primary teacher for and full partners in the education of their children and support their children's educational growth. (Section 1235(4) and 9101(20), ESEA; 20 U.S.C. 6381d(4) and 7801(20).) Section 1235(10) of the Even Start statute (20 U.S.C. 6381d(1) requires instructional components to be based on scientifically based reading research, to the extent that research is available.

Participant attendance: The actual number of hours that a participant attended classes or participated in program services.

Participant learning outcomes: Expected learning results for all participants. These outcomes include state-required performance indicators.

Participation calculations: Measurements of different program variables—program intensity, participant attendance, attendance rate, and length of enrollment. These calculations should be based on instructional activities within the four core Even Start components.

Participation expectations: Program requirements or policies for continued participation. The expectations state, for example, the number of hours of attendance or the percentage of time families are expected to attend program services.

Participation variable: An amount. Program intensity, participant attendance, attendance rate, and length of enrollment are variables in Even Start family participation patterns. The evaluator and program staff analyze variables to understand and explain differences in participants' achievement rates.

Partnership: An eligible entity that may apply for an Even Start subgrant is a partnership between one or more LEAs and one or more other non-profit public or private entities.

Percentile: A number showing how many cases, out of every hundred, fall below the point (score, amount) in question.

Performance indicators: Measurable indicators or standards established by states in response to the requirement to develop program quality indicators in three areas of adult learning achievement and three areas of child learning achievement; also called performance standards, state literacy indicators, or quality indicators.

Pie chart: A round (pie) chart that shows 100% of a variable divided into different parts; used to show the relationship of parts to a whole.

Pre-post testing: Giving the same test before a program intervention, and just after the intervention.

Professional development: As defined in section 9101(34) of the ESEA, specific activities that, among other things, improve and increase teachers' knowledge of the academic subjects the teachers teach, and enable teachers to become highly qualified. Professional development includes activities that are high quality, sustained, intensive, and classroom-focused in order to have a positive and lasting impact on classroom instruction and the teacher's performance in the classroom, and, activities that advance teacher understanding of effective instructional strategies that are based on scientifically based research. (Section 9101(34), ESEA; 20 U.S.C. 7801(34).)

Program director: The person(s) directly in charge of the local Even Start program; also called program coordinator, program leader, project manager, or program co-coordinator.

Program elements: As used in the Even Start statute, fifteen activities and services that local Even Start programs must provide (Section 1235, ESEA; 20 U.S.C. 6381d). These required program elements include: identifying and recruiting

families most in need of Even Start services; screening and preparing participants for full participation; developing flexible scheduling and support services; offering high-quality, intensive instructional programs in the four core instructional components; meeting specific staff qualifications; training staff; offering home-based instructional services; offering year-round services; coordinating with other programs; delivering instructional programs based on scientifically based reading research; supporting regular attendance and retention; creating reading readiness activities for preschool children based on scientifically based reading research; ensuring continuity of services; providing services to families most in need; and using an independent local evaluation for program improvement.

Program intensity: The total number of hours of instruction the average participant would receive if he or she were enrolled for the entire project year and had a 100% attendance rate.

Program theory: The underlying logic that explains how particular program activities and approaches are intended to produce short- and long-term results.

Qualitative methods: Evaluation approaches that are primarily descriptive and interpretative. Qualitative research typically uses observation, interviews, and document reviews to examine quality, meaning, and context.

Quantitative methods: Evaluation approaches that primarily involve numerical measurements and data analyses by statistical methods.

Range: All the values (amounts, numbers, scores) from lowest to highest; the distance a whole group covers.

Raw score: Basic units of observation; stated as the number correct, the number of errors, or the time for completion.

Relational database: A collection of data items organized as a set of formally-described tables from which data can be reassembled in many different ways without having to reorganize the database tables.

Reliability: A measure of stability, that is, the extent to which results will be the same each time a measure is repeated under the same conditions.

RFP or request for proposal: To be eligible to receive a new subgrant of Even Start funds, an eligible entity must submit an application to the state educational agency (SEA). The application must be in the form and contain or be accompanied by the information required by the SEA, and include specific information required by section 1237 of the Even Start statute. (Section 1237, ESEA; 20 U.S.C. 6381f.) Some states refer to the application as an RFA or request for applications.

Sample: A smaller part of a larger group of people (the whole group is “the universe”). The sample may or may not be chosen by chance. When used as a verb, sample means to choose this smaller group.

Scientifically based reading research (SBRR): Research that (a) applies rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties; and (b) includes research that –

- (i) employs systematic, empirical methods that draw on observation or experiment;
- (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;
- (iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations; and
- (iv) has been accepted by a peer-reviewed journal or approved panel of independent experts through a comparably rigorous, objective, and scientific review. (Sections 1208(6) and 1232(e)(4), ESEA; 20 U.S.C. 6368(6) and 6381a(e)(4).)

Significance: The degree to which a value is greater or smaller than would be expected by chance. Typically, a relationship is considered statistically significant when the probability of obtaining that result by chance is less than 5%.

Stakeholder: Individuals, groups, or organizations with an interest in how well a program functions and/or who have decision-making authority over program funding, implementation, and use of evaluation findings.

Standard score: A type of score derived from the standard deviation and the mean of a distribution of scores. Standard scores are superior to other types of derived scores because they can be added, subtracted, or averaged and facilitate comparisons across tests. Also called scale score.

Standardized: A consistent way of giving, scoring, and/or interpreting tests; the same or similar tasks are given under the same conditions to all students and scored the same way.

State coordinator: The state education official responsible for administering the federal Even Start grant to the state.

Support services: One of the 15 program elements that each local Even Start program must provide when those services are unavailable from other sources and are necessary for an eligible family's participation in Even Start activities. As indicated in the statute, examples of such support services include child care for the period that parents are involved in the Even Start program and transportation for the purpose of enabling parents and their children to participate in the Even Start program. (Section 1235(3)(B) and (C); 20 U.S.C. 6381d(3)(B) and (C).)

Technical report: Typically the end-of-year written report, it is the most informative of all reporting formats. It contains comprehensive analyses of participant outcomes, answers key questions, summarizes all evaluation findings, and presents recommendations for improvement.

Temporary Assistance for Needy Families (TANF): Federal block grant program that provides assistance and work opportunities to needy families by granting states federal funds and flexibility to develop and implement their own welfare programs.

Triangulation: Studying an issue using several different methods (e.g., a survey and focus groups), to determine whether or not results converge.

Universe: The full population in contrast to a sample, or a portion of a population group. An evaluator might give tests to all primary grade students in a district (the universe) rather than sample or select from the primary grade students.

Validity: The measure of how well a scale or test shows what it is supposed to show. There are several types of validity—each (concurrent, content, criterion, predictive) must be evaluated separately.