EERE Guide for Managing General Program Evaluation Studies

Getting the Information You Need

Prepared for the Office of Energy Efficiency and Renewable Energy (EERE)
Office of Planning, Budget and Analysis

Principal Author, Harley Barnes, Lockheed Martin Aspen
Lead Contractor, Gretchen Jordan, Sandia National Laboratories

February 2006

U.S. Department of Energy
Energy Efficiency and Renewable Energy
Bringing you a prosperous future where energy is clean, abundant, reliable, and affordable

Acknowledgments

This Guide for Managing General Program Evaluations (Guide) was completed for the U.S. Department of Energy (DOE) by Sandia National Laboratories, Albuquerque, New Mexico, USA under Contract DE-AC04-94AL8500. Sandia is operated by Sandia Corporation, a subsidiary of Lockheed Martin Corporation.

Gretchen Jordan, Sandia National Laboratories, was the lead contractor for the development of the Guide. Harley Barnes, Lockheed Martin Aspen, was the principal author. John Reed, Innvologie; and Martin Schweitzer, Oak Ridge National Laboratory made technical contributions. Michelle Clark, Lockheed Martin Aspen, assisted in the development of the graphics and prepared the Guide for publication.

Jeff Dowd, DOE's Office of Energy Efficiency and Renewable Energy (EERE), Office of Planning, Budget and Analysis (PBA), directed the work.

EERE reviewers were:

Christy Cooper, Hydrogen, Fuel Cells, and Infrastructure Technologies Program
Jean Diggs, Weatherization and Intergovernmental Program
Charles Hemmeline, Office of Planning, Budget and Analysis
Faith Lambert, Weatherization and Intergovernmental Program
Tien Nguyen, Office of Planning, Budget and Analysis

External peer reviewers were:

Edward Vine, Lawrence Berkeley National Laboratory
Marion Brown, Southern California Edison
Paul DeCotis, New York State Energy Research and Development Authority
Nick Hall, TecMarket Works
Mike Messenger, California Energy Commission
Iris Sulyma, BC Hydro
Dan Violette, Summit Blue Consulting
Carol White, National Grid USA Service Company

Notice

Table of Contents

Appendices

List of Figures

List of Tables

EXECUTIVE SUMMARY

More attention is being devoted to general program evaluations today than ten years ago. This is because of the increased emphasis in the last decade on good management practices in the Federal Government. This emphasis has stressed the importance of linking program resources with program performance.

General program evaluations include market needs assessments, process evaluations, retrospective outcome/impact assessments, and cost-benefit evaluations. These types of evaluation studies help managers like you determine if timely adjustments are needed in program design or implementation to improve the rate, or quality, of achievement relative to the committed resources. General program evaluations are in-depth studies of program performance and customer needs. They can be used to produce information about the linkage between program performance and resources and about how to improve performance. The benefits of conducting an evaluation are numerous. For example,

1) They can help you estimate how well your program is achieving its intended objectives
2) They help you improve your programs
3) They quantify results and cost-effectiveness, as necessary, to help you better communicate the value of your program.

This Guide:

1) Details and explains the five types of general program evaluations
2) Provides guidance on selecting the type of evaluation suited to the program to be evaluated, given the type of information required and budget limitations
3) Gives clear advice on managing the selected evaluation
4) Specifies important considerations in final report production and dissemination.

The Guide is intended for use by managers of both deployment and R&D programs within the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), although most of the examples of evaluations pertain to deployment programs.

EERE has published a separate guide for peer reviews—a specific form of process evaluation—that is a companion to this general program evaluation guide.[1]

The different types of general program evaluations usually require data-collection and analysis methodologies with which EERE's program managers may have little familiarity. It is not necessary to have in-depth familiarity with these methods to benefit from a general program evaluation, but program managers need to have enough familiarity to select and monitor an evaluation contractor who will make decisions about evaluation methodologies.

The Government Performance and Results Act of 1993 (GPRA), the recent White House President's Management Agenda, and Office of Management and Budget (OMB) circulars have established an expectation that credible evaluations of Federal programs be conducted on a periodic or ad hoc basis. This includes the programs implemented by EERE. Many program managers may not be familiar with the types of general program evaluations or how to initiate and manage them. This Guide takes managers through a high-level step-by-step process for designing and conducting general program evaluations.

[1] U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, "Peer Review Guide," August 2004

This Guide divides the program evaluation process into 14 easy-to-understand steps. Many of the 14 steps are illustrated with references to actual EERE program evaluations performed in recent years. The steps are as follows:

Planning and Formulating an Evaluation

Getting Started

Step 1: Decide the Evaluation Objectives
Step 2: Determine the Resources Available
Step 3: Determine the Timeline for Completing the Evaluation

Designing the Evaluation

Step 4: Construct or Confirm a Program Logic Model
Step 5: Specify the Questions the Evaluation Must Answer
Step 6: Develop a Research Design
- **6a:** Determine the Type(s) of Research Design Needed to Answer the Evaluation Questions
- **6b:** Select Data to Be Collected and Develop a Data Collection Plan
- **6c:** Select the Analytical Methods for Answering the Evaluation Questions

Step 7: Identify the Information that Will Go into the Evaluation Report
Step 8: Establish the Quality-Assurance Review Process
Step 9: Develop the Statement of Work and Select an Evaluation Contractor

Management During the Evaluation

Step 10: Develop an Evaluation Plan and Conduct the Evaluation Study
Step 11: Monitor the Evaluation Contractor During the Study

Disseminating and Applying the Results

Step 12: Determine Distribution of Final Report and Results and Distribute
Step 13: Make, or Monitor the Making of, Decisions about the Program Based on the Evaluation Results
Step 14: Establish/Update Program Records for Use in Future Evaluations

The steps are in the recommended order for consideration by a program manager who will manage a general program evaluation; however, there will be considerable interaction between the steps, and the manager can expect some of them to be revisited during performance of the evaluation.

As the Guide explains, program managers do not need to make all of the decisions described in these steps. An evaluation contractor will make the more technical decisions; however, the contractor will be guided by the program manager's specifications. Program managers can write more effective specifications and better judge which evaluation contractor's proposal will give the best evaluation for the money if they have a basic understanding of the methods used in designing and performing evaluations. This Guide provides information to help managers through the evaluation process.

1.0 INTRODUCTION

General program evaluations are systematic and objective studies, conducted periodically or on an ad hoc basis, to assess how well a program is achieving its intended goals. General program evaluation studies are management tools that answer a broader range of critical questions about program improvement and accountability than EERE's regular performance-monitoring and reporting activities.[2] They are performed as a matter of good management practice. These evaluations **have a retrospective focus**, with a view to assessing past performance and developing recommendations for improvements, with the exception of the evaluations of market needs that can have a current or prospective focus. General program evaluations usually require a level of detail in data collection and analytical methodology that goes beyond routine performance-monitoring reporting. (For example, they can help to explain why variations in performance reported by the Joule Performance Tracking System occur.)

General program evaluations help managers determine what kinds of timely adjustments may be needed in program design or implementation to improve the rate or quality of achievement relative to the committed resources.

Programs in the Office of Energy Efficiency and Renewable Energy (EERE) use several forms of evaluation to assess progress and to promote improvement:[3]

- ***Peer reviews*** by independent outside experts of both the program and subprogram portfolios to assess quality, productivity, and accomplishments; relevance of program success to EERE strategic and programmatic goals; and management.[4]
- ***General program evaluation studies*** by outside independent experts to examine process, quantify outcomes or impacts, identify market needs and baselines, or quantify cost-benefit measures as appropriate.
- ***Quarterly and annual assessments*** using DOE's Joule Performance Measurement Tracking System, R&D Investment Criteria, and the White House Office of Management and Budget's (OMB) Program Assessment Rating Tool (PART)[5]
- ***Technical Program Reviews*** by EERE senior management, Technical Teams, or Advisory Committees.

This Guide focuses on the management and use of general program evaluation studies performed by outside experts and contractors.

---

[2] Office of Management and Budget, "Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports," OMB Circular No. A-11 (2002), Part 6, Section 200.2.

[3] Technology validation and operation field measurement is also a form of evaluation used by EERE, but it is not covered in this Guide.

[4] EERE guidance for the peer review form of evaluation that uses independent outside experts is provided in a separate EERE Peer Review Guide. U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, "Peer Review Guide," August 2004.

[5] Separate guidance on performance monitoring and reporting is also provided by DOE's Chief Financial Officer (CFO) on the Joule Performance Measurement Tracking System and R&D Investment Criteria, and by OMB on PART. (See FY2006 Instructions for PART Worksheets (www.whitehouse.gov/omb/part/2006_part_guidance.pdf)).

Evaluators in all types of organizations recognize five types of general program evaluations. You will need to know them.

The five types of general program evaluations are:[6]

Process or Implementation Evaluations . . .

examine the efficiency and effectiveness of program implementation processes. The results of the evaluation help program managers decide how to improve program operations.

Outcome Evaluations (Retrospective Focus) . . .

estimate the success of outputs in achieving objectives. Findings show how well the program is achieving its intended outcomes in a specific time frame. This helps program managers decide on whether to continue the program as is, and at what level of effort.

Impact Evaluations (Retrospective Focus) . . .

take outcome evaluations one step further by estimating the proportion of the outcomes that are attributable to the program rather than to other influences. As with outcome evaluations, these findings help program managers decide whether to continue the program as is, and at what level of effort, but decisions based on impact evaluation findings can carry greater weight because they are based on outcomes that are likely not to have occurred without the program.

Cost-benefit Evaluations (Retrospective Focus) . . .

compare program benefits and costs. Cost-benefit evaluation shows the relationship between the value of the outcomes of a program and the costs incurred to achieve those benefits. The findings help allow program managers judge whether to retain, revise, or eliminate program elements.

Needs or Market Assessment Evaluations . . .

assess market baselines, customer needs, target markets, barriers to adoption of energy efficiency and renewable energy, and how best to address these issues by the program in question. Findings help managers decide who constitutes the program's key markets and clients and how to best serve the intended customers. When performed at the beginning of a program, needs and market assessment evaluations also establish baselines against which to compare future progress.[7]

[6] The first four types of evaluation are defined in a 1998 GAO report entitled "Performance Measurement and Evaluation: Definitions and Relationships." (www.gao.gov/special.pubs/gg98026.pdf). This report was reissued in hardcopy during 2005. The fifth type of evaluation was identified by other organizations and added.

[7] Market potential analyses that involve estimating efficiency or renewable-energy potential, including projections into the future, are not included in this definition.

A typical evaluation project will use more than one of these types of general program evaluation. Its content will depend on who needs the evaluation results and what they want to get out of them. This Guide will help you identify the correct type(s) of evaluation and evaluation methodologies to meet your needs.

1.1 Purpose and Scope

The purpose of this Guide is to help you create and manage ***objective, high quality, independent, and useful*** general program evaluations.[8] It has been developed for EERE staff and managers who have responsibility for planning, commissioning, and managing evaluation studies. The Guide and its step-by-step approach to managing general program evaluations can be used by those without prior training or experience in program evaluation. It should make it easier for them to take advantage of this useful and increasingly required program-management tool. (Appendix 1 summarizes the management steps.) Wherever possible, the steps are illustrated by examples drawn from evaluation studies already performed by EERE programs. The Guide is intended for use by managers of both deployment and R&D programs (although managers of R&D programs will find that most of the examples pertain to deployment programs).

This Guide will not answer all of the technical questions a program manager may have about evaluation methodology, but ***it will provide enough information to help program managers:***

1. Identify the questions that they need answered by a general program evaluation
2. Specify the type of evaluation(s) needed
3. Hire a qualified evaluation contractor
4. Monitor the evaluation's progress
5. Implement credible quality assurance (QA) controls
6. Ensure the evaluation report presents useful findings and recommendations
7. Ensure that the findings get to those who need them.

1.2 Why Perform an Evaluation?

More attention is being devoted to general program evaluations today than ten years ago. This is because the increased emphasis in the last decade on good management practices in the Federal Government has particularly stressed the importance of linking program resources with program performance. Good program-management practice is stressed in EERE's Strategic Plan.[9] One of the Plan's strategic objectives is to "Change the Way We Do Business" (to achieve excellence in business management). By using general program evaluations as tools for good management

[8] Many guides exist for planning and managing an evaluation. The steps in this Guide draw from several of them to create a planning guide specifically tailored to the needs of EERE program managers. The following are some of the evaluation-planning sources that were used in developing this Guide:
GAO, "Designing Evaluations," March 1991, GAO/PEMD 10.1.4 www.gao.gov/specdial.pubs/pc1014.pdf.
Ellen Taylor-Powell, Steele, S., and Douglah, M. "Planning a Program Evaluation," University of Wisconsin Extension, February 1996. www.cecommerce.uwex.edu/pdfs/G3658_1.pdf.
J. S. Wholey, Evaluation and Effective Public Management, Little, Brown and Company: Boston, 1983.

[9] http://www.eere.energy.gov/office_eere/pdfs/fy02_strategic_plan.pdf

practice, managers can improve their program to make it more cost effective while also contributing to EERE's other strategic objectives and to DOE's strategic goals and vision.[10] General program evaluations serve two critical purposes – program improvement and accountability. Many evaluations will be designed to serve both of these purposes.

- **Improvement:** General program evaluations help managers assess how well their programs are working by assessing the extent to which desired outcomes are being achieved and by identifying whether improvements are needed to increase effectiveness with respect to objectives. General program evaluation studies help program managers proactively optimize their programs' performance.

- **Accountability:** General Program evaluations also help program managers and others demonstrate internal and external accountability for the use of public resources. This includes fiscal responsibility, delivering services as promised in contracts, and individual performance plans.

Two major Federal directives in recent years have helped promote general program evaluations as tools for good management practice. The Government Performance and Results Act of 1992 (GPRA), as amended by the Reports Consolidation Act of 2000, requires each of the Executive Branch departments and many of its agencies to submit annual program performance reports to the OMB and to Congress.[11] General program evaluations provide important input to these reports.

In 2003, OMB took the performance evaluation process a step further by creating the Program Assessment Rating Tool (PART). As stated in the PART Guidance, "PART is a systematic method of assessing the

Examples

Using an Evaluation for Improvement Objectives.

An evaluation of the Industrial Assessment Center (IAC) Program recommended that:

- *The IACs put examples of their recommendations online for businesses to see and, possibly, act upon without audits. This could increase the energy savings from the program by leveraging assessments already performed.*
- *The program better understand the life cycle for the decision process to select energy efficiency investments and develop new program elements (e.g., executive training courses) to encourage acceptance of energy efficiency investments.*

[1]Oak Ridge National Laboratory, "Industrial Assessment Center Program Impact Evaluation." ORNL-CON-473, December 1999.

Using an Evaluation for Accountability Objectives.

A recent evaluation of the FEMP Energy Savings Performance Contracting (ESPC) program used a theoretical decision model to demonstrate that those agency managers who were aware of energy savings performance contracting moved through the decision process to implement energy efficiency improvements in much greater numbers than those who were not aware. This example is illustrated in greater detail in Section 3.3, Step 6a.

TecMRKT Works and Sandia National Laboratories, "2001 FEMP Customer Survey," June 2001.

[10] D. A. Beschen, "Some US Applied R&D Insights; President's Management Agenda & PART," 2004 WREN Workshop. www.wren-network.net/resources/2004eu/5.Darrell.Beschen-1.ppt

[11] See Office of Management and Budget, Circular A-11, Part 6, section 230, June 2002, for performance reporting requirements. This circular is updated annually with reporting instructions for each fiscal year; however, the June 2002 circular provides more detail than subsequent versions. To access the current Circular A-11, go to www.whitehouse.gov/omb/circulars/index.html.

performance of program activities across the Federal government. The PART is a diagnostic tool; the main objective of the PART review is to improve program performance."[12]

PART is divided into four sections: program purpose and design, strategic planning, program management, and program results/accountability. Each of these sections has a series of questions to assess whether a program has sound management practices and is producing results. General program evaluation studies can provide evidence to support a program's PART submission.

There are at least ten questions in the FY2006 PART that have direct or indirect relevance to evaluation. To the extent that objective and independent general program evaluation studies are completed, they can help to increase OMB's PART rating for a program, in addition to providing a program with ongoing feedback regarding improvements in program design and implementation. Appendix 2 lists the PART sections and their specific questions that have direct or indirect relevance for general program evaluations.

Examples

The importance of general program evaluations to the PART process is illustrated in two key elements of PART 2006:[1]

- *PART Question 2.6: "Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?*

  "A Yes answer would require regularly scheduled objective, high quality, independent evaluations that examine how well the program is accomplishing its mission and meeting its long-term goals. . . . Evaluations should be sufficiently rigorous . . . To be independent, non-biased parties with no conflict of interest would conduct the evaluation. . . . evaluations must be appropriate to the type of program. . . . R&D programs also should undergo independent reviews of relevance to their agencies, fields of science or technology, or customers in addition to assessing questions of performance. These reviews should conclude with reports documenting the findings and recommendations. . . ."

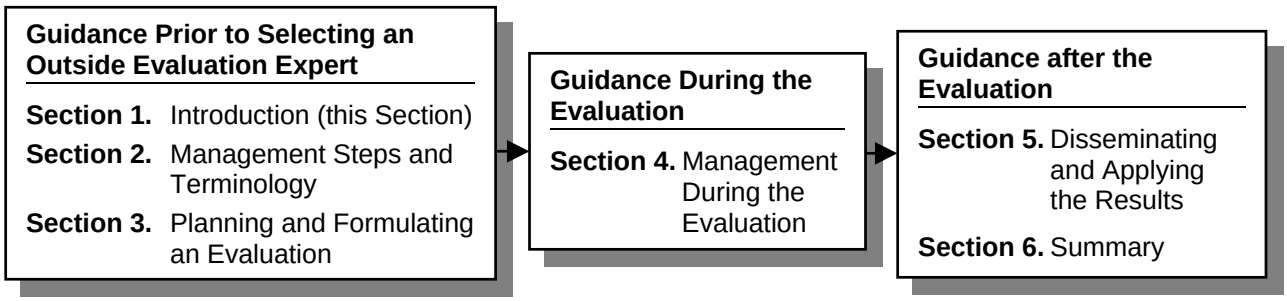
- *PART Question 4.5: "Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?"*

[1]OMB Budget Data Request No. 04-31, "Completing the Program Assessment Rating Tool (PART) for the FY2006 Review Process."

[12] FY2006 PART Guidance. The PART worksheet results contribute to OMB's comparative management and performance assessments across the Federal Government. See www.whitehouse.gov/omb/part/2006_part_guidance.pdf.

1.3 Guide Roadmap

The Guide is divided into six sections. Each part provides management guidance prior to selecting an outside evaluation expert, during the evaluation, and after the evaluation.



Appendices contain an evaluation management checklist, a model statement of work (SOW) for an evaluation study, EERE guidelines for evaluation quality assurance, a model report outline, and other useful information.

2.0 MANAGEMENT STEPS AND TERMINOLOGY

This part of the Guide:

- Lists the basic management steps for planning and managing general program evaluations.
- Defines a few of the more important terms that you may encounter when communicating with an outside evaluation expert. Appendix 3 contains other technical terms and definitions relevant to general program evaluations.

Even though program managers will solicit outside evaluation experts to perform their general program evaluation studies, they will need to clearly define and formulate the evaluation objectives and expectations before selecting a qualified contractor (see Steps 1 through 8 below). Therefore, all of the steps in this Guide will be important. The next section lists the management steps for conducting a general program evaluation. The remainder of the Guide describes them.

Familiarity with the steps in a typical general evaluation and with evaluation terminology will facilitate communication with the independent evaluation experts who perform the evaluation.

2.1 Steps in Planning and Managing a General Program Evaluation

The Guide divides the planning and managing process for a general program evaluation into 14 steps and describes briefly what each step entails. Steps 1 through 8 provide guidance for deciding what the evaluation should accomplish, and will help the program manager prepare to hire an outside evaluation expert (Planning and Formulating an Evaluation). Once hired, the evaluation expert will prepare a detailed evaluation plan (Step 10) that describes Steps 4 through 7 and what they will accomplish in greater technical detail. Steps 9 through 11 provide guidance for initiating and managing the evaluation (Management During the Evaluation). Steps 12 through 14 provide guidance on getting the maximum value out of an evaluation investment (Disseminating and Applying the Results).

The steps in this Guide appear in the order in which they are often performed in practice; however, it may be preferable to perform them, especially Steps 1 through 5, in a different order for specific evaluations.

The evaluation expert will develop the technical details of Steps 4 through 7 when the expert prepares the evaluation plan (Step 10). These steps appear in the order of management given in this Guide because *it is very important* that the EERE manager develop some idea about what Steps 4 through 7 will require before writing the SOW and selecting the expert (Step 9).

The steps are not prescriptive, but they *do* represent common practice for general evaluations. In that sense, it will be valuable to review this Guide in its entirety and become familiar with its concepts before beginning to plan and formulate an evaluation.

The following lists the steps:

Planning and Formulating an Evaluation

Getting Started

Step 1: Decide the Evaluation Objectives
Step 2: Determine the Resources Available
Step 3: Determine the Timeline for Completing the Evaluation

Designing the Evaluation

Step 4: Construct or Confirm a Program Logic Model
Step 5: Specify the Questions the Evaluation Must Answer
Step 6: Develop a Research Design
- **6a:** Determine the Type(s) of Research Design Needed to Answer the Evaluation Questions
- **6b:** Select Data to Be Collected and Develop a Data Collection Plan
- **6c:** Select the Analytical Methods for Answering the Evaluation Questions

Step 7: Identify the Information that Will Go into the Evaluation Report
Step 8: Establish the Quality-Assurance Review Process
Step 9: Develop the Statement of Work and Select an Evaluation Contractor

Management During the Evaluation

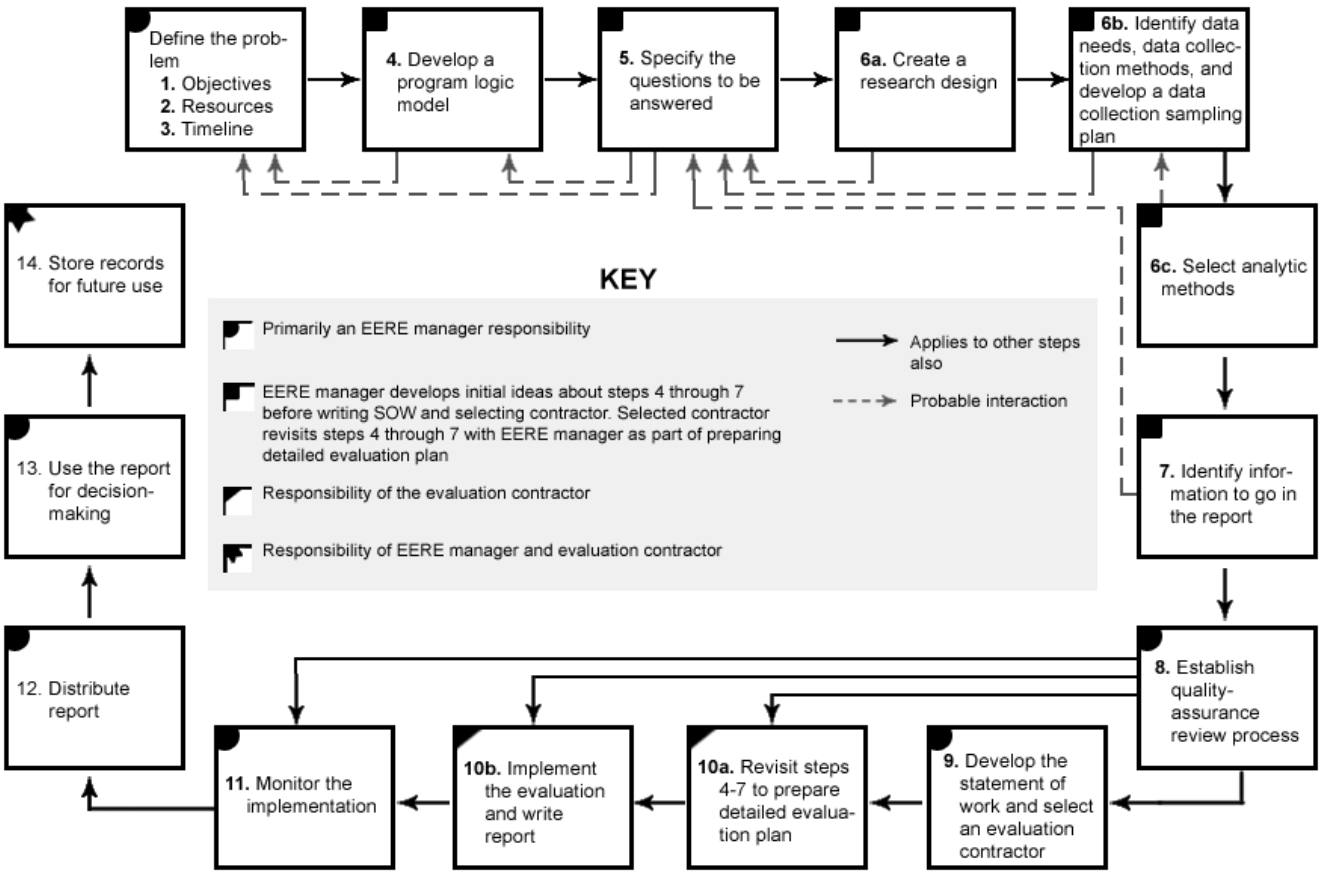
Step 10: Develop an Evaluation Plan and Conduct the Evaluation Study
Step 11: Monitor the Evaluation Contractor During the Study

Disseminating and Applying the Results

Step 12: Determine Distribution of Final Report and Results and Distribute
Step 13: Make, or Monitor the Making of, Decisions about the Program Based on the Evaluation Results
Step 14: Establish/Update Program Records for Use in Future Evaluations

Figure 1 shows the relationships between these steps. Keep in mind, however, that the circumstances of individual evaluations may recommend that they be performed in a different order, and that the EERE manager will use Steps 4 through 7 to prepare an SOW that delivers the information needed, but after selection, the evaluation contractor will develop the details of the work described in those steps.

Figure 1. Overview of Steps for Managing an Evaluation



2.2 Key Definitions and Terminology

Evaluation professionals use several terms that are common to the field of program evaluation to describe general program evaluation concepts. Knowledge of how these terms are used by evaluators will help program managers communicate their evaluation needs and expectations to an evaluation contractor. Several of the more important terms are defined in this section. Appendix 3 contains a more comprehensive glossary of terms used in general program evaluation planning and performance.

The terminology used in this Guide reflects usage by Federal evaluation experts. In a few cases it differs from that used by energy-program evaluation experts in the private sector. For example, the private sector has not adopted the distinction between "outcome" and "impact" used by logic modelers in this Guide. The private sector uses "gross" and "net" impacts (or outcomes) to describe the concepts intended by "outcome" and "impact" in this Guide. Appendix 3 attempts to bridge these differences. If you develop the impression while talking to an evaluation contractor that the terms you and the contractor are using may not have the same meaning, clarify them. This is critical to getting the information you need out of an evaluation study.

Independent Evaluation

For an evaluation to be independent, OMB requires that, ". . . non-biased parties with no conflict of interest should conduct the evaluation. Evaluations conducted by the program itself should generally not be considered 'independent;' however, if the

agency of program has contracted out the evaluation to a third party this may qualify as being sufficiently independent. Evaluations conducted by an agency's Inspector General or program-evaluation office might also be considered 'independent.'[13]

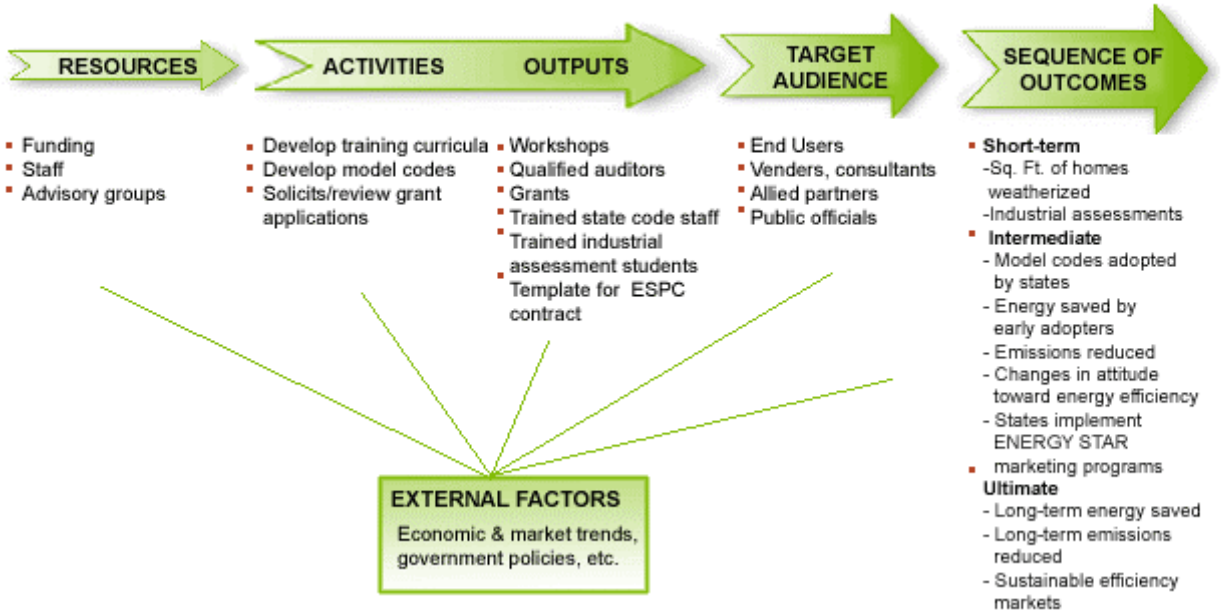
Cause, Causal, Causation, Attribution

No methodology exists that can perfectly account for all of the influences that may have produced ("caused") a result that a program is trying to achieve. Nonetheless, the terms, "cause," "causal," "causation," and "attribution" are often used in evaluation discourse to indicate that an evaluation has accounted for non-program (external) influences that may have contributed to the observed result that the program is trying to achieve and, therefore, some part of the observed result may be said to have been "caused" by the program. In this Guide, the words "cause," causal," "causation," and "attribution" are used in this sense.

Logic Model

A "logic model" is basic to planning EERE general program evaluations. A logic model diagrams the sequence of causes (resources, activities, and outputs) that are intended to produce the results (outcomes) sought by the program. It is common practice in general program evaluations to prepare a logic model of the program before moving too far into the evaluation process. Figure 2 shows a simplified illustration of a logic model.[14]

Figure 2. Simplified Program Logic Model for a Deployment Program



[13] Office of Management and Budget, "Completing the Program Assessment Rating Tool (PART)," March 2005, p. 29. www.whitehouse.gov/omb/part/fy2005/2005_guidance.pdf

[14] Step 5 provides additional guidance for creating a logic model for your program.

Logic model terminology is useful for promoting understanding of evaluation plans. Program managers use "inputs" (or resources) to design program "activities" which, in turn, produce the program "outputs" (e.g., energy audits) that managers hope will achieve the "outcomes" that the program's goals require (e.g., installations of energy conservation measures that lead to energy savings). (See Appendix 3 for formal definitions of these terms.) The important function of a logic model is to make you think about the logical relationships between the activities and outputs that produce the outcomes your program is designed to achieve. It will be seen in Step 5 that the point of using these terms is to focus evaluation thinking on all of the relationships in the program that contribute to its goals.

Impact

"Impact," as used in general program evaluation, refers to the portion of an outcome, e.g., energy savings or implementation of energy efficiency technologies, that it is estimated (by accepted methodologies) would *not* have occurred without the program. That is, it is the portion of the observed result that can be attributed to the program. It contrasts with the portion of an outcome that is attributed to other external influences such as economic or market trends.[15] The total amount of a targeted outcome that a program is trying to achieve is sometimes called the "gross" outcome. The amount that is attributed to the program is then called the "net" outcome. Cost-benefit evaluations use the results of impact evaluation rather than the results of outcome evaluations.

Indicator/Measure

An indicator is a variable that is measured or estimated to produce evaluation findings. It is also called a "performance indicator," a "performance measure," or a "performance metric." OMB defines a performance indicator as a particular characteristic that is used to measure an output or outcome.[16] "Square feet upgraded" and "computer tools distributed" are examples. Indicators must be measurable, but need not be observable. An example of a non-observable indicator is a measurement of whether a participant would have taken a particular energy-efficiency upgrade in the absence of the program based on the participant's own report.[17]

Defensibility

Defensibility is the ability of evaluation results to stand up to scientific criticism. It is described by another important set of terms—"validity," reliability," "strength," "accuracy," and "bias." Program managers will encounter these terms when considering alternative research designs (Step 6).[18] These are terms that describe whether the research design and the methods used to implement the evaluation can withstand technical review and critique by external experts and stakeholders (Step 8). This is an important issue because it determines whether the findings are acceptable to a peer review community. If the findings are eventually judged not defensible, they will have no value to those who expected to use them. If the purpose of the evaluation is to help make a decision based on the value of the program, the defensibility should be high.

---

[15] Office of Management and Budget, OMB Circular No. A-11 (July 2003), Part 6, Section 200.2.
[16] Office of Management and Budget, OMB Circular No. A-11 (July 2003), Part 6, Section 200.2.
[17] An evaluation of the Motor Challenge Program used such a "what would you have done" indicator to estimate outcomes that could be attributed to the program. Xenergy, "Final Report of the Motor Challenge Program," prepared for Oak Ridge National Laboratory, May 10, 2000, p. 3-11.
[18] The definitions provided are widely used in evaluation practice. The definition of "strength" is found in GAO, "Designing Evaluations,' GAO/PEMD-10.1.4, March 1991, pages 16-18. www.gao.gov/special.pubs/pe1014.pdf.

- ***Validity.*** This term is used to describe whether scientific standards have been applied to the research design to give it credibility. There are two types of research-design validity: internal and external validity. "Internal validity" refers to whether the research design minimizes the effect of non-program influences on the outcome so that claims of program influence are defensible. It is a concern principally with impact evaluations. "External validity" refers to whether one can claim that sample findings apply to the entire population from which the sample was drawn.

- ***Reliability.*** This term describes the degree to which, if you were to repeat the evaluation using the same design, indicators, and data-collection and analysis methods, you would get the same results.

- ***Strength.*** This is a general term that is used to describe the overall defensibility of the evaluation. GAO defines "strength" as: asking questions clearly; employing methods of analysis that are appropriate to the evaluation questions; supporting the findings with evidence; documenting the assumptions, procedures, and modes of analysis; and ruling out competing evidence of causation.[19] Step 8 lists many quality-assurance criteria such as those specified by GAO that can be used to assess strength. No scale of strength exists upon which all evaluators agree. Strength and weakness are judgmental terms, like "good" and "bad," and often a peer reviewer makes the judgment. A good evaluation contractor will be able to help with selecting methods that will produce as strong an evaluation as possible, given its budget.

- ***Accuracy.*** "Accuracy," also called "measurement accuracy," refers to the correspondence between the measurement made on an indicator and the true value of the indicator. In sampling, "accuracy" differs from "statistical precision." The latter describes the theoretical probability that the actual value of an evaluation indicator estimated by a sample lies within a specified range of the sample value. Accuracy describes the exactness of the measurements made in the sample. The sampling literature includes accuracy in the concept of "non-sampling error."

- ***Bias***. "Bias" refers to the extent to which a measurement method, a sampling method, or an analytic method systematically underestimates or overestimates a value. "Bias" is a concern of sampling, statistical precision, measurement accuracy, and validity.

Scope

The evaluation's "scope," as used in this Guide, is the types and amount of information, data collection, and analysis required of the evaluation.

NOTE:

You may also want to scan the glossary in Appendix 3 to familiarize yourself with some of the other terminology that is common to evaluation practice in and outside the Federal Government.

[19] GAO, "Designing Evaluations,' GAO/PEMD-10.1.4, March 1991, pages 16-18. www.gao.gov/special.pubs/pe1014.pdf.

3.0 PLANNING AND FORMULATING AN EVALUATION

This part of the Guide provides guidance for:

- Deciding when to perform a general program evaluation
- Getting started
- Designing an evaluation

General program evaluation studies are typically performed at program initiation and every 2-5 years during the period of program operation.

3.1 Deciding When to Perform a General Program Evaluation

There is no hard and fast rule on when to conduct a general program evaluation. Many factors can help to determine the need for one, e.g., budget constraints, PART reviews, changes in the program, and legislative requirements. It is always advisable to conduct a baseline evaluation while planning a new program. A follow-up general program evaluation should be conducted every two to five years thereafter to monitor progress. Table 1 indicates the times in a program's life cycle when it can benefit from a general program evaluation.

Table 1.
Guidance on the Timing of General Program Evaluations

Program Life Cycle Stage	Recommended Type of General Program Evaluation
Planning or early implementation stage	**Needs/Market Assessment:** before the program is initiated. (This evaluation will contribute to program effectiveness by determining the need for the program's services, identifying potential barriers to its prospective objectives, identifying the market segments that can most benefit from program services, identifying possible program partners, and if the program will be designed to change market behavior, establishing a market baseline (the existing behavior) against which to compare future behavior changes. If major changes are occurring in external factors that may affect program outcomes, e.g., energy prices and changes in minimum-efficiency standards, a baseline assessment may be needed every 4-7 years to support impact evaluations.)
During program operations	**Process evaluation:** once every 2-3 years, or whenever a need exists to assess the efficiency of the program's operations and barriers to its progress. Process evaluations can also be performed at any time to answer ad hoc questions regarding program operations. If results from consecutive process evaluations do not change, subsequent evaluations can be performed less frequently. **Outcome evaluation:** once every 2-3 years, or annually if results are to be used to support annual GPRA benefits analysis. Outcome evaluations can also be performed at any time to answer ad hoc questions regarding program outcomes. **Impact evaluation:** once every 3-5 years, or annually if results are to be used to support annual GPRA benefits analysis, or as often as cost-benefit analyses will be performed. *(continued)*

Program Life Cycle Stage	Recommended Type of General Program Evaluation
	Cost-benefit evaluation: once every 3-5 years. (Continued) **Needs/Market Assessment:** repeated as necessary to determine if there is a continuing need for the program's services, for identifying new market segments to target, or for updating the baseline in response to changes in external influences.
End of program	**Process and impact evaluations after the program has ended:** within one year of the end of the program. Apply process evaluation lessons to design of next-generation programs; use impact evaluation in a cost-benefit evaluation. **Cost-benefit evaluation:** within one year of the end of the program only if needed for accountability.

Because general program evaluations usually require six to twelve months (or more) to perform, a general program evaluation will not be conducted more often than annually. A less frequent schedule is common. Ideally a program office would have an evaluation strategy with a schedule of planned general program evaluations and with resources set aside for them.

3.2 Getting Started

To get started with planning and formulating a general program evaluation, you need to perform three steps:

- **Step 1:** Decide the Evaluation Objectives
- **Step 2:** Determine the Resources Available
- **Step 3:** Determine the Timeline for Completing the Evaluation

Step 1: Decide the Evaluation Objectives

In this step you will:

- **Determine the types of decisions about the program to which the evaluation will contribute**
- **Determine the evaluation's objectives**
- **Decide the type(s) of general program evaluation needed to satisfy the objectives.**

Steps 1, 2, and 3 will probably be performed concurrently. You may need to revisit Step 1 while you are designing the evaluation (Steps 4 through 8) because resource considerations (Step 2) affect all choices of evaluation method. Some program managers may wish to perform Step 2 first and assign roles and responsibilities before establishing the objectives of the evaluation or performing Steps 1 and 3. This choice would involve the key personnel in the setting of the evaluation objectives and its timeline. Regardless of the order in which you perform these first three steps, the objectives must be established before specifying the questions the evaluation must answer (Step 5).

Evaluation objectives are determined by careful consideration of the possible decisions to which the evaluation's results will contribute.

- **Programs use evaluations to help decide on future actions.** Ask what decisions about the program will be made as a result of the evaluation, who will make them, and when. Examples of the types of decisions made are:
  - Continue the program as is
  - Expand the program, consolidate components, or replicate components found to be most cost-effective
  - Reallocate funding within the program; add funding to the program; reduce program funding
  - Streamline, refine, redesign the program (e.g., to meet a pressing resource constraint)
  - Set more realistic objectives
  - Discontinue ineffective delivery components
  - Discontinue the program.

  The decisions listed above may not be based exclusively on the findings from a general program evaluation.

- **Specify the evaluation objectives.** The evaluation's objectives will specify what the evaluation must contribute in the way of findings to the overall decision process. Write the statements for the objectives. Figure 3 provides several examples of evaluation objectives, drawn from EERE and other general program evaluations.

- **Ask what kinds of information, in general, are needed to satisfy the objectives.** Table 2 provides examples.

- **Decide the type of evaluation needed to develop these types of information.** (The types of general program evaluation were defined in the Introduction.) Table 2 shows the kinds of information associated with each type of evaluation.

Figure 3. Examples of Statements of Evaluation Objectives

EVALUATION OF THE U.S. DEPARTMENT OF ENERGY MOTOR CHALLENGE PROGRAM

The primary objectives of this evaluation are to:

- Assess the effects of the Motor Challenge on the motor system purchase, management, and maintenance practices of end users who received tools, informational materials, or training services from the program
- Assess the effects of the Motor Challenge Program on the motor system specification and sales practices of vendors and consultants who received tools, informational materials, or training services from the program
- Assess the effects of the Motor Challenge program on utilities and government agencies that used Motor Challenge tools and materials to plan or implement their own motor system efficiency programs
- Develop a credible estimate of energy savings associated with improvements in motor systems due to changes in end-user, vendor, and utility practices, and programs attributable to Motor Challenge
- Place the program accomplishments mentioned above in the context of the larger market for industrial motor systems.
- Identify initiatives that are likely to enhance program results.

CALIFORNIA BUILDING OPERATOR CERTIFICATION AND TRAINING PROGRAM*

Evaluation objectives:

- Examine participants' satisfaction with program process and the content of training
- Gather participant and non-participant recommendations for enhancements to program process and content
- Understand how to better market the program to non-participants
- Document all participant post-program energy efficiency adoption actions.

DISTRIBUTED GENERATION INTERCONNECTION BARRIERS STUDY**

The objectives of this evaluation are to:

- Confirm that barriers do exist
- Provide illustrative examples
- Identify the kinds of barriers
- Quantify the barrier-related costs.

OTHER EXAMPLES

- Assess the impact of the program on customer awareness and knowledge
- Measure customer response to "follow-up" program elements designed to encourage audit participants to implement recommendations
- Examine program awareness, delivery channels, factors that influenced participation, program effects and customer satisfaction.
- Document energy efficiency actions taken by program participants compared to actions taken by non-participants
- Estimate energy savings accruing from participation in the program over time; verify the reported energy savings results of the program
- Determine if there have been any changes in the building characteristics of program participants between fiscal years
- Evaluate the effectiveness of program modifications made in a specific fiscal year
- Complete a customer segmentation analysis of the primary target population
- Explore the barriers to participation in program activities and develop recommendations for improving the promotion and targeting of existing services, as well as new programs and services.

* Pacific Gas & Electric Company, "Evaluation of the 2003 Statewide Building Operator Certification and Training Program," prepared by Research Into Action, Inc. March 2005.

** National Renewable Research Laboratory, "Making Connections, Case studies of Interconnection Barriers and their Impact on Distributed Power Projects," NREL/SR-200-28053, May 2000.

Table 2.
Illustrations of the Types of Information Associated with Different Types of General Program Evaluation

If You Need These Types of Information	Perform This Type of Evaluation
Customer needs and target markets, e.g., the extent of the populations that can be helped, whether gaps exist in the coverage of these populations or in the services they need, characterization of market baseline, whether new technologies or efficiency practices should be added to the program, barriers to participation.	**Needs or market assessment**
Efficiency of program implementation processes, e.g., effectiveness of specific activities, what works and what does not work, where additional resources could be leveraged, participant satisfaction.	**Process evaluation**
Quantified outcomes of the program and judgments of success of the outputs in achieving the results intended, e.g., energy savings achieved, non-energy outcomes, market effects (e.g., changes in retailer stocking practices), and any results that were not intended. These findings are sometimes called "gross impacts."	**Outcome evaluation**
Quantified outcomes that can be attributed to the program's outputs, i.e., outcomes that would not have occurred without the influence of the program. These findings are also sometimes called "net impacts."	**Impact evaluation**
Quantified measures of performance relative to funding, e.g., the relationship of program benefits to costs, or of a single activity's benefits to its cost. Benefits may be defined as gross or net benefits; however, net benefits will carry more weight in expert review of the evaluation. Benefits are usually quantified in dollars.	**Cost-benefit analyses**

Depending on the objectives and amount of resources available, a program may decide to perform more than one type of evaluation in the same time frame.

Step 4 calls for program managers to establish a logic model for the program to be evaluated. However, managers may want to develop the model as part of Step 1 to help them and their team define the decisions about the program that must be made and to identify the objectives of the evaluation.

Step 2: Determine the Resources Available

In this step you will determine the resources available and how these resources will determine the content of the general program evaluation.

You may need to revisit this step while you are designing the evaluation because resources affect choice of evaluation method. In any case, you must begin the evaluation design process with a sense of the resources available.

Deciding on the resources that will be committed to a general program evaluation involves simultaneous consideration of (1) the importance of the program decisions to which the evaluation will contribute, (2) the resources needed to satisfy the evaluation's objectives, and (3) the resources that the program can afford. The importance (to the program) of the information that the evaluation will develop should strongly influence the resources committed. The more important information needs will require more defensible evaluation methods, and more defensible methods usually cost more. *(See Step 6.)*

Resources consist of budget, staff, and time.

Budget:

- Some state, electric, and gas utility organizations have used a percent-of-annual-program-cost rule of thumb to set the cost for general program evaluations. These have ranged from 1% to 15%. However, such rules of thumb assume a number of factors that are not always made clear.
- As noted above, many factors influence evaluation cost. Broadly speaking, however, three factors weigh more heavily on cost. These are:
    1) The type of evaluation (described in Section 1.0 and Table 2)
    2) The scope of the information requirement, e.g., number of questions and size of the sample or census (discussed in Section 3.3, Step 5)
    3) The defensibility required of the information results (described in Section 2.2 and Step 8).

Table 3 shows illustrative ranges of cost for general program evaluations for these and other factors. The actual costs may differ from those shown: a larger evaluation scope will cost more, while a smaller scope will cost less. Measurement methods with higher accuracy, e.g., metering, can increase the cost. Most of the factors that affect cost are discussed in the steps for evaluation design (Steps 4 though 8).

Staff:

- Establish key internal staff roles and responsibilities, such as deciding whether available EERE staff has the time or skills to take a part in the evaluation itself, and how big a part EERE staff will take in assisting the contractor with data collection.
- Decide who will handle the administrative management of the evaluation.

Table 3.
Illustrative Costs for General Program Evaluation Studies

Type of Evaluation with Illustrative Scope	Defensibility		Other Factors Influencing Cost
	Lower*	Higher*	
Needs or Market Assessment Illustrative Scope: opportunities to serve a wider customer market	$50,000 - $70,000	$70,000 - $150,000 (can be more)	• Difficulty in identifying and contacting the target population • Choice of survey methodology (e.g., on-site, mail) • Number of customer segments • Need for OMB clearance
Process Evaluation Illustrative Scope: customer satisfaction measurement	$20,000 - $40,000	$30,000 - $100,000	• Difficulty in identifying and contacting the target population • Choice of survey methodology (e.g., telephone, mail, Web) • Need for OMB clearance
Outcome Evaluation Illustrative Scope: quantification of 5-8 direct and indirect outcomes (also referred to as "gross impacts")	$50,000 - $100,000	$70,000 - $200,000	• Number and complexity of outcomes • Source of information (e.g., survey vs. program records) • Availability of a program-implementation baseline • Analytical methodology used to estimate outcomes • Need for OMB clearance
Impact Evaluation Illustrative Scope: quantification of 5-8 direct and indirect outcomes attributable to program (also referred to as "net impacts")	$150,000 - $200,000	$200,000 - $500,000	• Number and complexity of outcomes (scope) • Source of information (e.g., participant and non-participant surveys) • Availability of a program-implementation baseline • Research design used to control for outside influences (e.g., control group vs. participant self-report of program influence) • Methodology used to estimate net outcomes
Cost-benefit Evaluation Illustrative Scope: Quantification of energy, cost, and environmental benefits and costs from one or more elements of the program	$40,000-$80,000	$60,000 - $100,000	• The difficulty in quantifying the benefits. • The costs of the program that produced the benefits (Recognizing partner and customer costs as well). • Extent of confounding influences between expenditures and benefits.

* Cost ranges are a function of scope. The ranges shown are based on discussions with energy-program evaluation experts.

Time Constraints (if any):

- Determine when information from the evaluation is required for the intended decision making. Can you wait six months, 12, months, or even 18 months before receiving the findings and recommendations from an evaluation study?
- Time constraints (if any) will also influence Step 3 (Timeline for Completing the Evaluation).

Available Program Records and Data (if any):

- The cost of constructing a database of the output and outcome history of the program can be significant. If program output and outcome data have been collected and recorded in a useable database from the beginning of the program, the cost of an evaluation may be reduced significantly. Programs are encouraged to develop and maintain a record-keeping database of routinely collected participant and program information for use in current and future evaluation studies.

Step 3: Determine the Timeline for Completing the Evaluation

In this step you will determine the timeline, schedule, and key milestones for conducting the general program evaluation.

This timeline is the timeline for the entire evaluation process, from determination of the objectives to making the decisions that will be based on the evaluation results (Step 1 through Step 14). The time line for performing the evaluation itself (Step 11) is a **part** of this overall timeline.

Work backwards from the end of the evaluation process.

- Determine when the information from the evaluation is needed.
- Allow time for quality assurance review of the results (see Step 8).
- Estimate the time it will take to perform the evaluation. If the evaluation is likely to require a survey to collect data, allow time for the OMB to approve the survey.[20] Step 6b and Appendix 5 contain guidance on obtaining OMB clearance to conduct a survey.
- Determine when the evaluation must begin in order to deliver its information when it is needed.
- Account for the administration time required to hire an evaluation expert.

By now, you have established the objectives, the kinds of information required for decision-making, the available resources, and the timeline. You can now begin to think about how the evaluation will be conducted, i.e., the *design* of the evaluation. In the course of this process, you may need to revisit Step 2 (resources available) and Step 3 (timeline).

[20] Surveys of Federal Government employees about Federal Government activities do not require OMB clearance.

3.3 Designing the Evaluation

The program manager does not need to make all of the decisions required for the evaluation's design. An evaluation contractor will make the more technical decisions; however, the evaluation contractor will take its cues for these decisions from the specifications in the request for technical services. The program manager will need to have a basic understanding of the methods used in designing and performing evaluations in order to give those cues. The program manager will also need such an understanding in order to judge which evaluation contractor will give the best evaluation for the money. To help, this section provides a brief introduction to the steps in designing an evaluation.

As you proceed through this section, you will need to keep in mind that the parts of an evaluation research design are interrelated. Ultimately, you will need a written management document—the Evaluation Plan—that lays out the plan for performing the evaluation. When working with an evaluation contractor, *require that the contractor develop such an Evaluation Plan* before the actual evaluation gets underway. (See Step 10 for more details about the Evaluation Plan).

Step 4: Construct or Confirm a Program Logic Model

Step 4 will describe how logic models can help you design a more effective evaluation by:

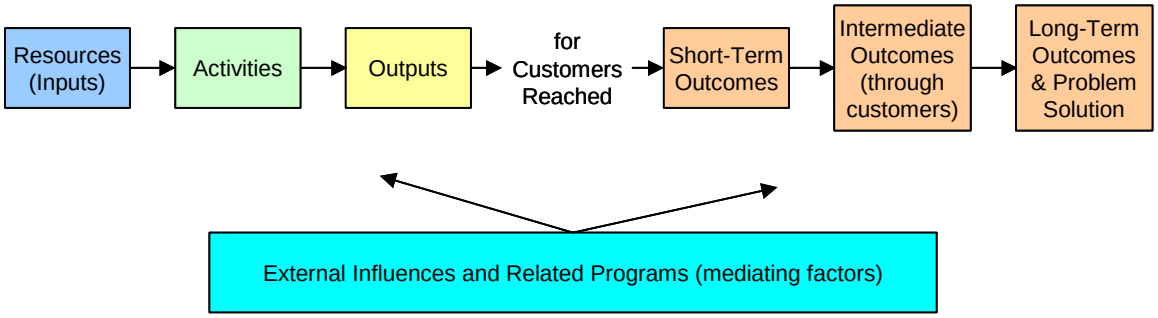
- **Explicitly describing the problem to be solved** by the program and making explicit the logical (theoretical) linkages between the program's activities, outputs, and outcomes that can solve the problem;
- **Identifying performance indicators** for which data may be collected by evaluation studies or routine monitoring;
- **Developing and prioritizing questions** related to the evaluation objectives;
- **Helping to identify alternative explanations** for projected outcomes;
- **Identifying activities** that are similar across multiple programs so that common indicators can be developed to enhance the efficiency of the general program evaluation process.

Construction of a logic model is not essential to designing an evaluation study, but it is highly recommended.

Before you can evaluate a program, you need to have a good description of the program so that you know what is included in the scope of the evaluation. A program's design reflects an underlying theory (the program theory) about how and why the program's activities and outputs will achieve the intended program outcomes. A logic model diagrams the design and helps to

make the theory explicit so that it can be examined more carefully.[21] Figure 4 illustrates the basic elements of a program logic model. A well-designed logic model requires you to think about what it is your program is trying to accomplish and how its activities produce outputs that will, in turn, produce the outcomes the program is trying to achieve.

Figure 4. The Basic Elements of a Logic Model



A logic model focuses thinking on the links between resources and performance. The logic model diagram, and the group process to define it, does this by organizing the program's elements into a framework of plausible cause-and-effect relationships. The framework comprises, in logical order, resources (sometimes termed inputs), activities, outputs, and outcomes (direct and indirect short-term, mid-term, and long-term outcomes). Feedback loops can also be added to show the relationships among logic elements.

Program managers and evaluators of a program sometimes want to see more detail themselves but show a simplified model to other managers. For this reason, multiple levels of logic models may be developed; however, the evaluation effort should not be diverted into an expensive discussion of the level of detail, especially the activity details For evaluation purposes, the model should include the appropriate level of detail needed to examine the relationships of interest.

Logic models may be in table form or a diagram with text.

Examination of the underlying program theory can help reveal the logic of these relationships, and thereby the logic of the program. Asking questions about the external, non-program factors that might influence an expected outcome can show non-program relationships and suggest alternative hypotheses about how outcomes occur. The process of examining the underlying theory, making explicit the logical relationships between the program components, and asking about external influences can suggest the need for changes to the program design or the evaluation plan.

As with any model, a logic model is a simplification of what's going on. If the program can be effectively diagramed as a flow of processes from resources to outcomes, a linear model of the program is appropriate. When activities or outputs interact with other program activities or

[21] A useful discussion of logic models, including a brief description of the history of logic models applied to program evaluation and a stage-by-stage process for constructing them, can be found in, McLaughlin, J. A., and G. B. Jordan, "Logic Models: A Tool for Telling Your Program's Performance Story," *Evaluation and Program Planning*, Vol. 22, No. 1, February 1999. Copies of this article are available from EERE's Office of Planning, Budget Formulation & Analysis. The University of Wisconsin–Extension Website also has useful resources on the development of logic models. www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html.

outputs, two dimensions may be needed to adequately capture the process that the evaluation needs to consider. If a program is made up of several subprograms, it may be necessary to diagram the subprograms as well as the program. In this case, the model of the program and one or more of its subprograms is said to be a "nested" logic model. Such a nested model would be needed to identify the program's dependent relationships with the activities and outputs of its subprograms. It can also be used to identify common evaluation questions that cut across the subprograms and program, thereby improving evaluation efficiency.

Figure 5 illustrates a logic model constructed at a high level (less detail) for a deployment program (DOE's Building Energy Codes Program). It illustrates the types of relationships that can be identified to trace the effects of program inputs to expected outcomes.

A logic model makes it easier to identify researchable evaluation questions that support the evaluation's objectives. Evaluators can ask, "What indicators represent what has been happening in the program that matches the boxes in the logic model?" and, "What questions should be asked to quantify the indicators?" Questions can also be identified to evaluate the efficiency of the relationships between the program elements. For example, the logic in Figure 5 suggests outcome evaluation questions about the extent to which states are training on code-compliance tools (an indicator of an outcome) and process evaluation questions about the efficiency of developing code revisions (relationship between activity and output). Identification of alternative hypotheses about why outcomes occur points the way to researchable questions about the potential effects of external factors.

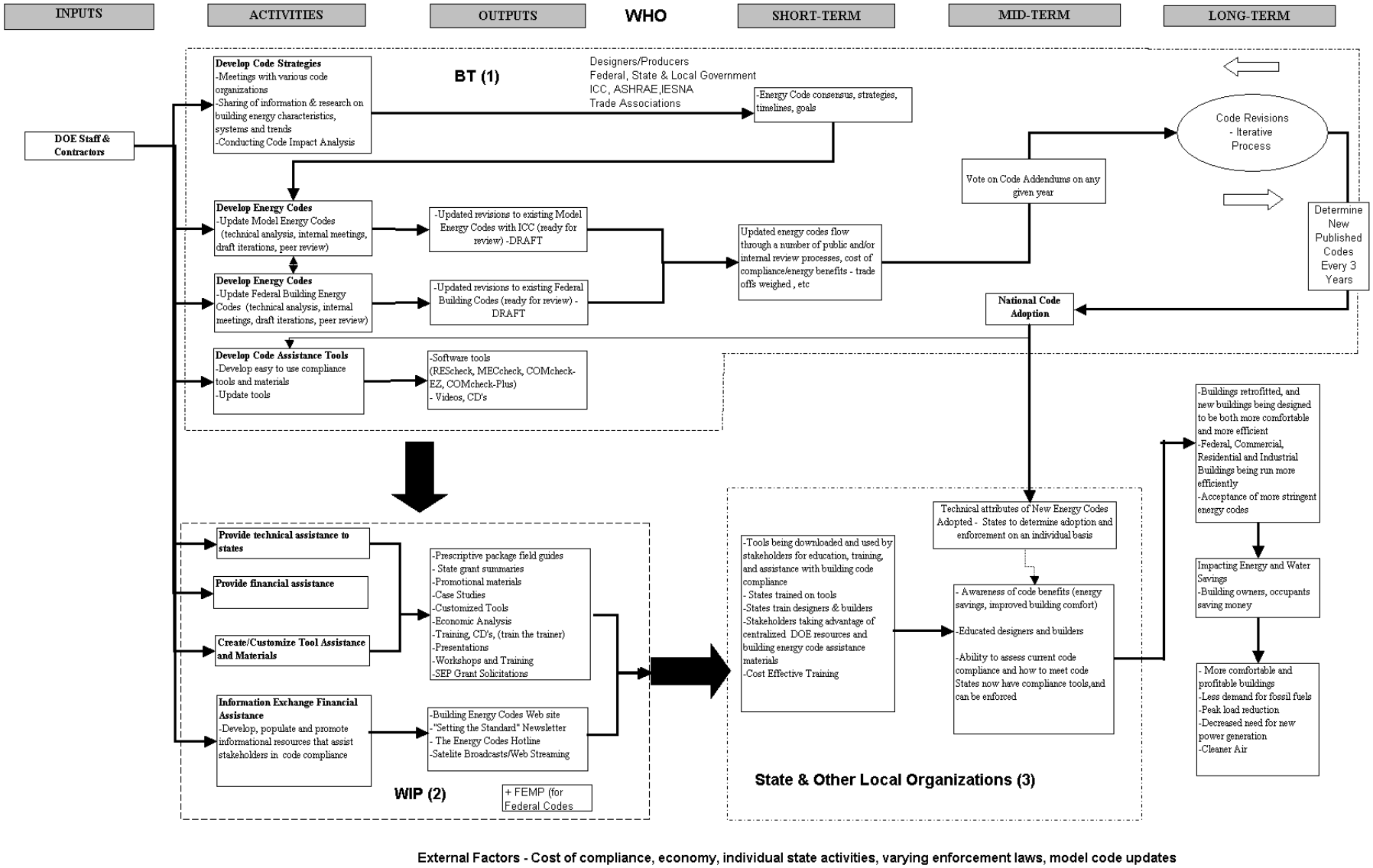
The discipline of developing a logic model of the program can help identify unexpected benefits of the program. For example, development of a logic model requires that evaluators understand all of the possible ways the program's short-term outcomes might ripple through the program-targeted population to produce the types of outcomes desired by the program. Some of these ripple effects occur when people replicate desired actions without involvement in the program or after earlier participation in the program. Once such additional outcomes are identified, evaluators will know to ask questions about if and why they occurred. Without the investigation of program theory stimulated by a logic model exercise, however, some of these potential outcomes may go unnoticed and uncredited.

Example: Identifying and Quantifying Indirect Outcomes

In 2002 a general outcome evaluation of the Industrial Assessment Center (IAC) Program devoted a special effort to identifying the program's indirect outcomes. The effort began with a consideration of all of the possible "ripple" effects that the outputs (industrial assessments, student training, and literature) might produce. The result led to a number of hitherto unconsidered indirect outcomes and increased the program's energy savings outcome by 25% over the savings due to direct outcomes alone. Figure 6 shows, from left to right, the energy savings adjustments due to the following types of indirect outcomes: spin off, internal replication of the recommended upgrades, external replication of the upgrades, delayed implementation (ultimate outcomes), decommissioning of the upgraded facility (an 18% loss in savings), and the net increase in the estimates of the program's energy-savings outcome.

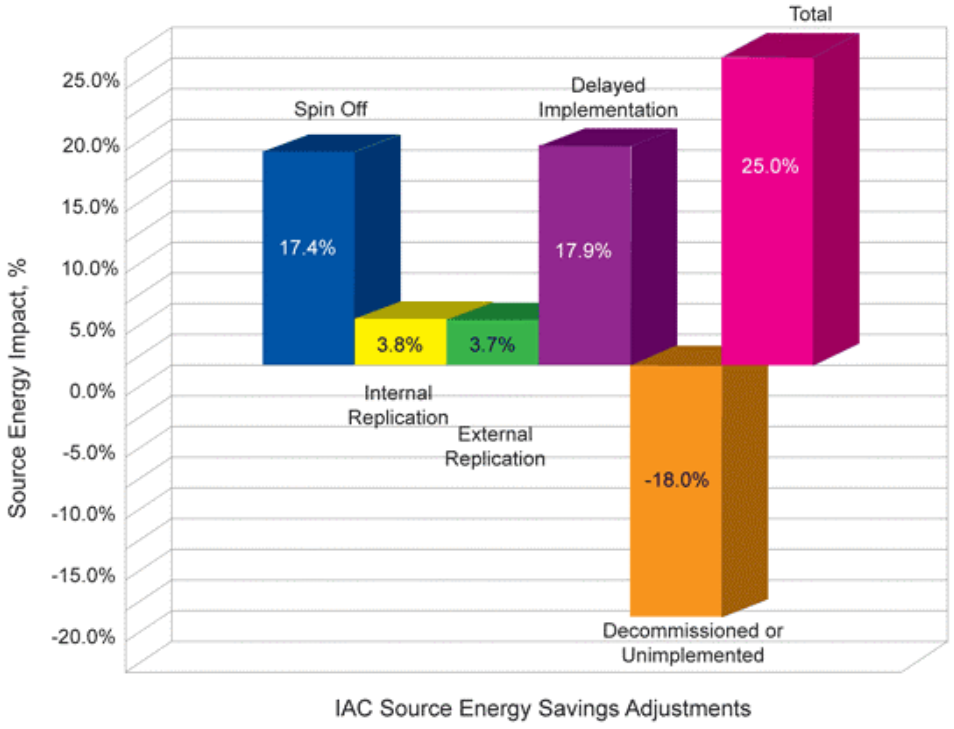
Oak Ridge National Laboratory, "Industrial Assessment Center Program Impact Evaluation, ORNL/CON-473, December 1999.

Figure 5. Building Energy Codes Program High-Level Logic Model



Abbreviations:
BT: Building Technologies Program
WIP: Weatherization & Intergovernmental Programs
(1): Numbers in parentheses represent sequence of actions

Figure 6. Illustration of Energy Savings Benefits from Indirect IAC Program Outcomes



*(**Source:** ORNL, Presentation on the Industrial Assessment Center Program Impact Evaluation)*

Sometimes, too many questions are of interest to program managers. When this occurs, the logic model helps to identify the questions that are most directly related to the evaluation's objectives. These questions are given higher priority to ensure that they will be asked during the evaluation. A logic model makes it easier to make and get agreement on these sometimes-difficult budget-related evaluation choices.

A logic model should not be static. As the program matures, its logic model should be revisited at least annually to check the assumptions embedded in its theory and to update it for lessons learned and changes in its external environment. Such periodic revisits can help managers identify where program adjustments will most effectively enhance program outcomes.

Note:

Later in this Guide you will see that every possible question emerging from a logic model need not be asked. You will need to prioritize the questions for investigation. However, having such a model will give you confidence that you are identifying the most important questions.

Logic models are valuable tools, but like tools, they should be a means, not an end in themselves. The diversion of effort from meeting the evaluation's principal objectives to building and developing consensus on a detailed logic model has been a source of criticism. In some instances the expense of logic modeling is justified because the effort can avoid costly premature impact evaluation. In other cases it is justified by benefits to the program external to evaluation, such as team building and identification of information essential for program design or redesign. The program can avoid the diversion-of-effort problem by (1) defining and keeping to a work plan that budgets for both developing

program logic and completing the evaluation, and (2) by going into only as much detail as is needed for the evaluation task at hand.

A logic model can be prepared for programs during the planning stage or at any time after implementation. Further information on logic models can be found on EERE's evaluation Web site.

In summary, logic models can help:

- **Make explicit the problem addressed and the theory of change underlying the program's inputs, activities, outputs, and outcomes (both direct and indirect)**
- **Identify key program relationships on which to collect data that can be used to monitor performance routinely and conduct general program evaluations**
- **Develop and prioritize questions for the evaluation to answer.**

You have now established the objectives of the evaluation (Step 1), estimated the resources available and timeline for the evaluation (Steps 2 and 3), and developed a program logic model for your program (Step 4). You are now ready to focus the evaluation by using the logic model to help you select the questions the evaluation must answer to meet its intended objectives.

Step 5: Specify the Questions the Evaluation Must Answer

In this step you will use the logic model and/or evaluation objectives to:

- **Help clarify the reasons** why you want to undertake an evaluation, e.g., to make particular decisions about the program
- **Identify higher-level, more general questions** the evaluation must answer and prioritize them
- **Select specific, researchable questions** the evaluation must answer.

This is a very important step. The evaluation must ask questions whose answers will satisfy the evaluation objectives. If the answers to the questions do not satisfy the objectives, the evaluation resources will be wasted. To develop relevant questions, it will be useful to start with a table that arranges the following from left to right:

1) **What type of evaluation is needed to develop the needed information?** What kind of information is needed to inform decisions about the program? (See Step 1)
2) **What are the evaluation objectives?** The objectives should be appropriate to the decisions under consideration and information needed. (See Step 1)
3) **Specify and prioritize the higher-level general questions that need to be answered to satisfy the evaluation objectives.** Determine what information is most needed and when. Define the general questions to provide this information. These prioritized

general questions are used to begin the process of specifying more detailed questions that will supply the required information. Some questions may not be answered if it is decided that the more important questions require all of the current evaluation resources. A multi-year evaluation strategy is helpful in that it can schedule coverage of all the general questions over a period of time.

4) **Select specific, researchable questions that can be asked to answer the general questions.** The most important questions should receive enough resources to develop defensible answers. Ensure that questions about outcomes include both direct and indirect outcomes. Pose all of the questions that you think are relevant. You will screen them later.[22]

The choice of more specific questions is very important.

1. They must be related to the evaluation objectives, and they must be researchable.
2. Your evaluation contractor will address the questions that you ask in the evaluation study, so be sure they are the right ones.
3. If you ask the contractor to propose the questions, make sure you are satisfied with the questions proposed.

Table 4 illustrates how a table may be constructed and used to help identify the questions and data that will be needed for an evaluation of an EERE deployment program. The examples in Table 4 are taken from *Draft Plans and FY 2001 Task Description: FEMP-Wide Evaluation and Performance Monitoring.* A more exhaustive list of illustrative questions may be found in Appendix 4.

[22] (GAO, "Designing Evaluations," GAO/PEMD-10.1.4, March 1991, pp. 12-14. www.gao.gov/special.pubs/pe1014.pdf)

Table 4.
Evaluation Type, Evaluation Objectives, General Evaluation Questions, Priorities, and Specific Research Questions

1 Type of Evaluation	2 Evaluation Objectives	3 General Evaluation Questions	4 Priorities High Medium Low[23]	5 More Specific Researchable Questions
Needs/Market Assessment	Identify how FEMP can accelerate the efficiency with which Federal agencies use energy. Identify which Federal agencies most need FEMP assistance. Create a baseline for future evaluations.	Q1: What is preventing Federal agencies from giving energy efficiency improvement a higher priority for annual funding? Q2: What do Federal agencies need that FEMP can provide to increase the number of efficiency upgrades they implement? Q3: What are the Federal agencies most in need of FEMP assistance that have not accepted it? Q4: What energy efficiency measures have been installed and/or what is the level of energy use prior to participation in FEMP's program?	Q1 High Q2 High Q3 Medium Q4 High	What are the market and agency barriers to adopting better energy and water management technologies? How are FEMP actions directly and indirectly meeting specific customers needs or lowering barriers to action? What customers is FEMP serving? Which need its services most? What is the energy use per square foot of an office building prior to participation?
Process Evaluation	Assess the adequacy of program funding relative to objectives. Determine if funding is being used as intended. Determine if populations that can benefit from the program are being served well. Identify opportunities to improve effectiveness of activities and outputs	Q1: What level of overall investment in energy efficiency are we leveraging with our spending? Q2: Do the Federal agencies perceive that we are helping them meet their energy-efficiency upgrade goals? Q3: How can we make our services more productive for Federal agencies?	Q1 High Q2 High Q3 High	How much does FEMP spend and on what activities? What is the total investment in energy and water projects? Are FEMP partnerships leveraging funds and capabilities? What is the quality of FEMP service and products? What can FEMP do to improve its services and its service delivery, generally and specifically (e.g., Web site) Is FEMP reaching the right customers and are they satisfied?

[23] The priority levels shown are for illustrative purposes only.

1 Type of Evaluation	2 Evaluation Objectives	3 General Evaluation Questions	4 Priorities High Medium Low[23]	5 More Specific Researchable Questions
Outcome Evaluation	Quantify the achievements of program outputs and outcomes against planned time frame. Assess whether further outcomes are possible and how to achieve them.	Q1: Have overall energy savings by the Federal Government increased from year to year? Q2: How many quads of energy savings are in the pipeline? Q3: Is progress toward energy-efficiency upgrade goals, by agency satisfactory? Can they meet these goals? Q4: Are there any actions possible, by agency, that have not been undertaken?	Q1 High Q2 High Q3 Medium Q4 Low	Is FEMP making progress, as indicated by FY2002-2005 measures such as percent participation in the Procurement Challenge; or savings identified in audits, demonstrations, and projects in the pipeline? Is the Federal Government, by agency, on track to meet its goals? What agency actions/projects (retrofit, procurement) are possible, or in the pipeline, demonstrating that those goals are being met?
Impact Evaluation	Assess the net effect of the program's activities, i.e., the proportion of the outcomes that can be attributed to the program instead of to other influences.	Q1: How much were they achieving before each of FEMP's initiatives? Q2: How much of the quantified outcome can the program claim? Q3: Which FEMP initiatives helped more than others? Q4: What would have caused Federal agencies to invest in energy efficiency upgrades had FEMP's programs not existed?	Q1 High Q2 High Q3 Medium Q4 Low	To what degree did a FEMP program contribute to specific benefits? Which FEMP tools helped more than others to create the benefits? What external factors would have caused agencies to create savings without FEMP?
Cost-Benefit Analysis	Determine program cost-effectiveness Determine the cost-effectiveness of individual outcomes, outputs, or goals, where possible.	Q1: Are the benefits from FEMP actions greater than the total of FEMP and customer costs?	Q1 High	What are the energy savings and emissions reductions attributable to FEMP initiatives, as determined by an impact evaluation? What are the savings to the taxpayer as a result of FEMP initiatives? What are FEMP's costs associated with the quantified benefits?

The following example shows several specific evaluation questions used in the evaluation of another program, the former DOE Motor Challenge Program.

Example: Selection of Specific Evaluation Questions

The following are examples of specific questions that were asked by an EERE evaluation:

- *How many end-user facilities, vendors, utilities, and government agencies received materials, tools, and training services from the various Motor Challenge components?*
- *To what extent did Motor Challenge participants adopt the "Best Practices" recommended and supported by the program's tools, informational materials, and training services?"*
- *What portion of the reported changes in motor system practices was attributable to Motor Challenge?"*
- *How much energy did changes in motor system practices attributable to the Motor Challenge program save?"*

Xenergy, Inc., "Final Report, Evaluation of the Motor Challenge Program," prepared for Oak Ridge National Laboratory, May 10, 2000, chapter 1.

Step 6: Develop a Research Design

In this step you and your evaluation contractor will select the data collection methods, analytical methods, and timing of the data collection needed to develop defensible answers to the questions selected in Step 5. You may need to revisit Step 2 and review the available resources before leaving Step 6.

The evaluation research design is the research strategy that permits defensible findings to be deduced from the evaluation data. It consists of:

- The questions and indicators for which data will be collected
- Inventory of existing data and identification of data gaps
- The method and timing by which the data will be collected
- The populations from which the data will be collected
- The choices of research accuracy, sampling precision and confidence level, and degree of defensibility for the results

- The method of analysis used to produce the evaluation results
- The method of reasoning from the results to answers to the questions.[24]

The research design will be documented in the Evaluation Plan described in Step 10.

Step 6 has several sub-steps. Although they are described below as discrete sub steps, in practice, they must be performed concurrently because decisions made in one directly affect the decisions made in the others. Research design, data collection, and analysis are separate but interrelated activities.

For example, an analysis method such as a case study may use several methods of data collection, and the research design establishes how the findings of a case study will provide defensible answers to the evaluation questions. Trade-offs on questions, research design, data-collection methods, analytic methods, and defensibility always occur. The evaluation expert should propose a research design as part of the proposal process and finalize and document the design in the Evaluation Plan.

Step 6a: Determine the Type(s) of Research Design Needed to Answer the Evaluation Questions

The evaluation contractor must propose a logical approach to inferring answers to the evaluation questions from the data collected by the evaluation study. This logical approach plus the data-collection method(s) and analytical method(s) constitutes the "research design." This step discusses one form of research design, the experimental design needed for impact evaluations.

The evaluation contractor's proposed approach to inferring answers to the evaluation questions from the collected data can vary from simply tabulating the findings of a customer satisfaction survey to inferring the net outcome of the program from the results of an experimental design. Methods such as tabulating descriptive measurements and finding the statistical significance of a relationship between variables are usually not thought of as research designs, but, in fact, the process of going from the results of these analytical procedures to answers to evaluation questions involves logic and, therefore, implies a research design. However, these common methods are not discussed further here because the logical process involved in using them to find answers to questions is relatively straightforward. They are mentioned because it is important to understand how the evaluation

[24] Two reasonably easy-to-understand publications are available that can help you pull together all of the pieces needed to design an evaluation:

The GAO report, "Designing Evaluations" (GAO/PEMD-10.1.4, March 1991), especially chapters 2 and 3 (www.gao.gov/special.pubs/pe1014.pdf), provides guidance in selecting a research design for different evaluation objectives. Its terminology is slightly different than that used here, but the guidance is relevant.

The electric and gas utilities of California have prepared a framework that helps energy-efficiency evaluators choose research elements for program evaluations in general. The framework (500-plus pages) is available at www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf.

contractor proposes to derive answers to the evaluation questions from the data collected. Tabulation and the relationships between variables receive additional discussion under analysis methods in Step 6c. This discussion in Step 6a is about the special type of research design often used for impact evaluations, i.e., experimental designs.

If you need to determine the proportion of a quantified outcome that can be attributed to the program instead of to external influences, i.e., you need to conduct an **impact evaluation**, then some form of credible research design is required that will enable the study to infer this proportion. This design should be able to estimate what actions participants would have taken (outcomes) had your program not existed. The difference between what participants would have done and what they actually did is the amount of the observed outcome that you can attribute to your program. Evaluation research designs that allow you to make such claims of effect are called "experimental" or "quasi-experimental" designs.

Experimental Designs

Experimental and quasi-experimental designs are data-collection and analysis strategies that use deductive reasoning to estimate whether a program's outcomes can be attributed to the program's activities and outputs or whether they were likely to have occurred anyway. "Experimental" designs use randomly assigned participant and control groups with before-and-after measurement. They are the "gold standard" of evaluation research. OMB guidance for implementing PART calls these research designs "randomized control trials" (RCT), and recommends such designs as the most appropriate type to demonstrate the effectiveness of Federal programs.[25]

However, they are rarely used in energy program evaluations because they require random assignment of the target population to participant and control (non-participant) groups. This is not possible with programs whose success depends upon voluntary participation.[26] In such cases, group assignments are often made using non-random methods, e.g., by matching non-participants to participants on key characteristics. When matching is used rather than random assignment to select a non-participant comparison group, the term "control group" is sometimes replaced with "comparison group," and the designs are often called "quasi-experimental" designs.

Quasi-Experimental Designs

When the gold standard cannot be used because of the voluntary nature of a program, an approximation to it must be used. Designs using comparison groups offer such approximations.

[25] Office of Management and Budget, "What Constitutes Strong Evidence of a Program's Effectiveness?", p. 1. www.whitehouse.gov/omb/part/2004_program_eval_pdf.

[26] When participants voluntarily participate in a program, a type of bias called "self-selection bias" enters into the results. This bias alludes to the probability that the participants have a predisposition to be interested in the program's intervention, e.g., have a prior interest in energy efficiency. This creates two issues for the validity of the evaluation results: (1) it is expensive to find *non*-participants with similar predispositions for a comparison group, and (2) even if they were identified, the results could not be generalized to the broader population group that the program might be targeting because the part of the population without this predisposition would not be included in either group. When this issue is relevant, it is usually acknowledged as a source of unknown bias.

There are a variety of research designs using comparison groups; here are three of the more popular:[27]

1. **Before-After Comparison Group Design:** Compare program-participants and non-program-participants on before-program and after-program measurements. The amount that the program participants changed their behavior compared to the amount the non-participants changed is the amount the program caused. For example, how did the before-and-after efficiency-upgrade actions in the school system of a Rebuild America community partner compare to the before-and-after efficiency-upgrade actions taken, if any, in the school system of a non-Rebuild America community with similar characteristics?

2. **After-Only Comparison Group Design:** A less defensible variant of this design eliminates the before-program measurements and simply compares the two measurements at the same point after the participants participated in the program. The program effect is deduced by comparing the activities of the two groups.

3. **Before-After Participant Group Time-Series Design:** If you do not have a good non-participant comparison group, compare trends in participant behavior before and after participation. For example, if a pattern of low weatherization activity is seen before participation and the amount of activity suddenly jumps at about the time of program participation and continues at a higher level, conclude that the difference is due to the program. This design has less defensibility than the two described above, but costs less to implement and may be all that is feasible with the available data.

Table 5 illustrates the use of a theoretical decision model with a before-after comparison group design to attribute the numbers of Federal energy-efficiency outcomes (projects) to awareness of FEMP's energy-savings programs.

It uses one program, the energy-savings performance-contracting (ESPC) program, in its research design. The decision model borrows from diffusion theory to assume the various stages that agency decision-makers will go through to make an efficiency-upgrade decision.[28] Then, it demonstrates that Federal installations that received help from FEMP in use of the ESPC (participants) had implemented more energy-efficiency projects than installations that did not work with FEMP (non-participants) (over the same period of time) (compare columns (3) and (7)). Further, it also shows that agencies that were familiar with ESPCs but not receiving FEMP assistance with them *had not moved as far along* the decision path toward implementing ESPCs as those who did receive the FEMP assistance. (Compare column (6) to column (3)).

[27] The names for these designs have been adapted from D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally & Co., 1966.
[28] E. M. Rogers, *Diffusion of Innovations*. New York: Free Press, 2003

Table 5.
Example: Use of a Theoretical Decision Model with a Before-After Comparison Group Design

		Percent of FEMP ESPC participants (N=101)			Percent of non-participants in the FEMP ESPC Program who are aware of it (N=101)			
		(1) Before hearing about FEMP	(2) After involve-ment with FEMP	(3) Movement from (1) to (2) within stage	(4) Before hearing about FEMP	(5) After hearing abut FEMP	(6) Movement from (4) to (5) within stage	(7) Percent of unaware non-participants (N=188)
Stage of Decision	Unaware of ESPC	24	0	-24	21	0	-21	63
	Aware of ESPC	27	10	-17	31	40	+9	24
	Persuasion	12	7	-5	9	10	+1	5
	DecisionNo ESPC	10	7	-3	14	16	+2	1
	DecisionYes ESPC	3	21	+18	8	9	+1	1
	Implementation of Yes decision	7	24	+17	4	6	+2	4
	Confirmation of decision	18	32	+14	14	18	+4	2

A useful discussion of comparison groups in the context of Federal program evaluations and PART may be found in, Office of Management and Budget, "What Constitutes Strong Evidence of a Program's Effectiveness?"[29] It is recommended that this document be referenced in any statement of work used to solicit evaluation expertise.

Non-Experimental Designs

Research designs that do not use control or comparison groups are considered to be "non-experimental" designs.[30] The following two non-experimental research designs are also used for impact evaluations, although they have very weak defensibility. These designs do not use valid or reliable data collection methods to account for any of the possible external influences that might have caused the observed difference. Their only advantage lies in their low cost in comparison to experimental research.

4. **Participant Group Before-After Design:** Quantify participant behavior one time before (baseline) and one time after participation. Conclude that any difference is a result of the program. The possible effects of external influences may be acknowledged by hypothesizing their existence or by asking the participants whether they would have taken the action without the program (see design #5).

5. **Participant Group Self-Report Design:** Quantify participant behavior one time after participation and ask participants to tell you whether they would have taken the measured

[29] www.whitehouse.gov/omb/part/2004_program_eval.pdf
[30] Office of Management and Budget, "What Constitutes Evidence of a Program's Effectiveness?", p. 3. www.whitehouse.gov/omb/part/2004_program_eval.pdf. The second definition of non-experimental design given in the OMB document, "indirect analysis" using an independent panel of experts, is more appropriate for R&D projects.

actions had the program not existed. Participants may also be able to tell the researcher which external influences affected their actions; however, any effort to quantify these external effects will lack credibility. Similarly, the defensibility of participant claims about what they would have done, or energy they would have used, if they had not participated is weak. Some respondents will have a tendency to give the interviewer what they think is a socially acceptable answer or an answer that will make them look good; therefore, this research design also has weak defensibility. Nevertheless, this design is sometimes used for impact evaluations of relatively small energy programs both within and outside of the Federal Government because it is relatively inexpensive and does not depend on a pre-program baseline measurement.

The following is an EERE example of the use of a participant-group self-report design.

Example: Use of Participant Group Self-Report Design

An evaluation of the Motor Challenge Program asked respondents who reported using MM+ [MotorMaster software] to guide replace — versus — repair decisions whether "the motors repaired in the previous year would have been greater, less, or about the same if the software package had not been available."

Xenergy, "Final Report, Evaluation of the Motor Challenge Program," prepared for Oak Ridge National Laboratory, May 10, 2000, p. 3-11.

Table 6 will help you select a type of design to use for an impact evaluation:

Table 6.
Selection of Evaluation Research Design for Impact Evaluations

Level of Defensibility Needed	Funding Resources Needed	Comparison Group Before-After History Likely to Be Available	Participant Before-After History Likely to Be Available	Relevant Design Type
Highest	High	Yes	Yes	Before-After Comparison Group
Higher	Medium	After-only	After-only	After-Only Comparison Group
Medium	Medium - lower	No	Yes, and a time series of measurements is available	Before-After Participant Group Time Series
Low	Low	No	Yes	Participant Group Before-After Design
Lowest	Low	No	After-only	Participant Group Self-Report

The design you select will influence your data collection options in Step 6b.

There are several types of experimental and quasi-experimental designs. Determining which is best for different evaluation findings is beyond the scope of this guide. If you have not had prior training in experimental research design, but believe you need to conduct an impact evaluation, it is recommended that you seek expert assistance in assessing the options, or leave the choice of approach to the evaluation expert(s) who propose(s) the evaluation.

A good introduction is found in chapter 3 of GAO's "Designing Evaluations," (GAO/PEMD-10.1.14, March 1991) www.gao.gov/special.pubs/pe1014.pdf.

The relationship of research design to PART requirements is discussed in www.whitehouse.gov/omb/part/2004_program_eval.pdf.

A more technical, but understandable and short overview is, D. T. Campbell and J. C. Stanley, Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Co., 1966. The names and descriptions for the first three experimental designs above are based on names and descriptions in this overview.

Step 6b: Select Data to Be Collected and Develop a Data Collection Plan

In this step you will use the specific evaluation questions that you developed during Step 5 to:

- **Identify indicators by which to quantify answers to the questions**
- **Identify sources for the indicator data**
- **Identify alternative methods for collecting the indicator data**
- **Comply with OMB clearance requirements for performing a survey**
- **Decide on the content needed for a data collection plan.**

You can expect to gain from this step an overview of the more common methods for gathering the data for an evaluation and guidance for determining which will be useful for your evaluation.

Data collection is the process of taking measurements on the indicators that will be used to answer the specific questions identified in Step 5.[31] The program manager may choose to leave most of the decision making for data collection to an evaluation expert; however, a basic understanding of the commonly used alternatives will help the manager evaluate the recommendations offered.

[31] "Data collection" includes any activity that produces information that can be used to answer evaluation questions, e.g., surveys, informal interviews, focus groups, and compilation of data from program records.

What to Collect: Choice of Indicators

Indicators are the metrics, or researchable variables, for which the evaluation must collect, or develop, data. In Step 5, you selected and prioritized the specific questions to which the evaluation must provide answers. These were the questions needed to satisfy the evaluation's objectives. Under Step 6b, indicators must be selected that will provide the answers to these questions. You will want to be sensitive to defensibility and cost when you select indicators. Table 7 illustrates three indicator choices that were available for estimating how often recipients of MotorMaster software used the software to decide whether to repair or replace a motor. Indicator #2 in Table 7 is the choice actually used for the evaluation.[32]

Table 7.
Example: Illustration of Indicator Alternatives for a Specific Evaluation Question about Energy Savings

Specific Evaluation Question	Possible Indicators	Defensibility	Cost to Collect	Other Considerations
How often did recipients of MotorMaster software use it to make a decision to repair or replace a motor in the last year?	1. Answer to a question about number of times that recipients of MotorMaster software used it to make a decision to repair or replace a motor in the last year.	Low: Recall of frequency of an activity is often overstated.	Low	N.A.
	2. Answer to a question about whether recipients of MotorMaster software used it to make a decision to repair or replace a motor in the last year "always," "most of the time," "half of the time," "less than half," "hardly ever." (Indicator actually used by evaluator.)	Medium: Recall of frequency in terms of "more" or "less" scales is often more accurate.	Low	Defensibility declines if used to estimate quantitative results
	3. Ask recipients of MotorMaster software to keep a log of the number of times they use the software to make a choice between repairing or replacing a motor and count record of choices.	Highest of the three alternatives: frequency can be independently counted.	High. OMB clearance probably required.	Difficult to get recipient cooperation.

Developed from Xenergy, "Final Report of the Motor Challenge Program," prepared for Oak Ridge National Laboratory, May 10, 2000, p. 3-11.

[32] Xenergy, "Final Report of the Motor Challenge Program," prepared for Oak Ridge National Laboratory, May 10, 2000, p. 3-11. The alternatives shown are not mentioned in the evaluation report, but they are typical of the choices the evaluators would have faced.

Sources of Data

Data are referred to as "primary" if they are collected specifically for an evaluation (or other purpose) and "secondary" if they are collected by another project for another purpose. The Guide discusses secondary data sources first. Examples are:

- EIA energy end-use data
- Census data
- Energy savings coefficients, i.e., estimates of energy savings (e.g., kilowatt hours) per unit outcome (e.g., installation of an efficient motor), that were developed for one EERE program and may have relevance to another.

If applicable secondary data are available, it is advisable to use them to supplement routinely collected data and other primary data because secondary data will significantly reduce data-collection costs. However, two very important caveats must be considered. The data must be relevant and their transfer to the program for which they will be used must be defensible. See Appendix 9 on Lessons Learned for Improving the Quality of EERE Evaluation Studies. In particular, if energy-savings coefficients or gross or net estimates are available from the evaluation of another program, the program manager must ensure that the circumstances under which they were developed and the method of developing them are appropriate to the purposes of the current program's evaluation. Among other considerations, an energy-savings estimate would need to fit end-user industry and size profiles, as well as the application profile, to be credibly applied to other end-user populations and technology applications. If the secondary data do not satisfy such considerations, their use will not be defensible, and they should not be used for the current program.

Alternative Data-Collection Methods

A variety of methodological options exist for collecting data on (measuring) the indicators. Table 8 lists the more common methods, along with several of their key characteristics.

Table 8.
Comparisons of Data-Collection Methods

Data Collection Method	Relative Defensibility	Relative Cost for a Given Survey Size	Comments
Surveys	Varies	Varies	In this table and Guide, the term "surveys" means the collection of data in accordance with generally accepted methods that support statistical inferences, e.g., random sampling for the selection of respondents. The table uses the term "interviews" for data collection without statistical requirements. Both terms presume the use of data collection instruments designed in accordance with generally accepted principles of valid and reliable data measurement. *(continued)*

Data Collection Method	Relative Defensibility	Relative Cost for a Given Survey Size	Comments
In-person surveys	High	Usually high	Confidence in the accuracy of the measurements is usually highest. In the case of measurements of subjective data such as opinions, however, the selection and training of the interviewers is critically important to accurate measurement.
On-site metering, use of other types of measuring equipment	High	High	Energy-use metering is sometimes used in outcome and impact evaluations and often used in evaluating the energy savings for specific buildings. Industry-developed guidance for the use of metering for such purposes is published in the International Performance Measurement and Verification Protocol(s) (www.ipmvp.org) and for Federal buildings in FEMP's M&V Guidelines (http://www.eere.energy.gov/femp/financing/superespcs_mvresources.cfm)
Building simulation modeling	Medium-to-high	High	Whole-building simulation is more often used to assess alternative building configurations relative to an energy use goal or to diagnose compliance with efficiency standards. Occasionally such information may be input to general program evaluations.
Utility billing data	High, provided the evaluator understands how to get the desired information out of utility billing files	Low	Energy-use histories for specific energy customers of energy utilities may have relevance for evaluations, e.g., an evaluation of the Weatherization Assistance Program. Typically the request is made to the utility by a third party such as a state energy office. The utility must agree to provide the data voluntarily, and privacy issues may be involved. It can be challenging to understand and process utility billing files.
Mail surveys	Medium	Medium, usually higher than telephone	Non-response is an issue, although methods exist for compensating for non-response. (See multi-mode methods below.) The accuracy of responses about objects (e.g., recognition of logos) with whose names the respondent may not be familiar can be improved by providing a picture or drawing.
Telephone surveys	Medium	Medium	Non-response is an issue, although methods exist for compensating. (See multi-mode methods below.) Telephone interviews usually take less time to complete than the other methods because you have more control over the rate of response. The validity of responses to complex questions is a serious issue. If call lists must be purchased from a list vendor, the evaluation contractor will need to provide the vendor with documentation that the lists will be used for research purposes to avoid conflict with the National Do Not Call Registry. https://www.donotcall.gov/FAQ/FAQDefault.aspx

(Continued)

Data Collection Method	Relative Defensibility	Relative Cost for a Given Survey Size	Comments
Web site or e-mail surveys	Medium to Low	Low	The principal source of weakness is obtaining a probability sample so that statistical precision can be claimed. If many members of the population of interest do not have access to computers it is difficult to claim a probability sample. Many Web or e-mail surveyors demonstrate that their respondents represent the overall population and make the claim. This method is growing in popularity; however, care must be taken to demonstrate that the population that is capable of being sampled is the population in which you are interested.
Interviews	Low	Medium	As used here, "interview" means the collection of data through protocols that will not support statistical inference. These are informal one-on-one question-and-answer sessions, usually with small numbers of respondents, which are designed to gather insights from experts on particular topics. The can be conducted in-person, by telephone, or e-mail. See also "Focus groups."
Focus groups	Can make a defensible contribution to process and needs evaluations, but otherwise low	Low	Focus groups are used to probe selected respondents in-depth for their reasons for a choice or their opinions regarding an event or object of interest. The findings of focus groups do not have statistical precision because the samples are very small (8-12 persons) and are almost always non-probability samples.
Observation, e.g., mystery shopping	Can make a defensible contribution to process evaluations, but otherwise low	Low The cost of mystery shopping can increase to medium if travel for national-level research is required.	Mystery shopping is used to estimate participants' adherence to program rules without the participants' being aware of being evaluated. Usually the samples are non-probability samples. Observations are usually used as tools in process evaluations.
Literature review	Depends on purpose to which put, low to medium	Low	Literature reviews may contribute to *meta-evaluations* (borrowing the results from evaluations of comparable programs that operate under similar circumstances, synthesizing the findings, and applying them to your program) or obtaining anecdotal information, e.g., to use as evidence of external influences. Literature review may also be used to expand one's knowledge about the latest program theory for the purpose of developing effective evaluation questions.
Program records and reporting	Depends on purpose to which put, low to high	Often lowest	Program records and reports often serve as sources of data for indicators. As such, they may be the most accurate data available, which can contribute to the evaluation's strength. If the accuracy of their data is questionable, however, it can weaken the evaluation.
Multi-mode methods (use of more than one method for a single data-collection)	Usually higher than the individual methods used alone	Higher than the individual methods used alone; however, synergy may help to reduce cost.	Combinations of mail, telephone, and in-person data collection for a particular survey can increase response rate and help to evaluate bias due to non-response. OMB sometimes expects the use of multi-mode methods for these purposes.

Another data-collection choice involves whether the evaluation collects data (1) from the entire population of participants (like a census), or (2) from a sample of the population. Either option may be used for any type of evaluation; however, like most of the other choices, the choice has implications for cost and defensibility of the results. Table 9 highlights these options.

Table 9.
Options for Selecting the Number of Respondents from Which to Collect Data

Option	How Many Are Measured and Resulting Statistical Precision of Estimates	Rank Order of Contribution to Defensibility*	Relative Cost
Census	**Measure entire population.** Statistical precision is not applicable because you are counting every outcome and, therefore, have a full rather than partial enumeration.	Highest	Usually Highest
Sample Probability sample: Simple random and stratified random Systematic	**Measure a subset of the population.** Probability of a unit entering the sample is known. Sampling precision depends on the number of items, e.g., participants, measured. The more measured, the better the precision.		Medium. The cost will increase with the sample size.
Any non-random method of sampling	**Measure a non-randomly selected subset of the population.** Probability of selection unknown. Statistical precision, not applicable. Carefully selected representative samples are sometimes claimed to have properties "similar to" probability samples.	Lowest	Usually lowest

* Each of the defensibility judgments shown in Table 9 assumes that the data collected by the option are measured accurately. Inaccurate measurements will damage a claim of defensibility made by any data-collection or analysis method.

It will be very useful when communicating with evaluation experts to be aware of the difference between "statistical precision" and "accuracy" as used in survey-based data-collection activities. "Statistical precision," also known as "sampling error," applies to samples and consists of two parts: (1) how close (within a plus or minus interval) you want a sample estimate to be to the true population value, and (2) the probability of getting a sample whose results will lie inside the desired interval. The former is the width of the interval within which you want the true value of the variable being estimated to lie in relation to the estimated value, e.g., plus or minus 10%. The probability of getting a result that will lie inside this interval is the "confidence level" that the sample will deliver a result within this interval. Usually, "statistical precision" and "confidence level" together are specified as a "confidence interval,"

A common issue associated with taking a census and sampling is non-response, i.e., the fact that you will not be able to obtain information from some members of the population selected for your survey (unit non-response) or that those who respond do not answer all of your questions (item non-response). Non-response threatens the strength of the results. The usual method easing this threat is to require the evaluation contractor to demonstrate that those in the census or sample who did not respond to a survey are similar to those who did.

e.g., +/-10% with 90% confidence, or often, 90 +/-10%. If statistical results are desired for any of the specific questions, a program manager may ask the evaluation contractor to recommend the target confidence interval(s) for the findings.

"Accuracy" refers to the correspondence between the measurement made on an indicator and the true value of the indicator. Accuracy describes the exactness of the measurements made in the sample. In the sampling literature, accuracy is part of the concept of "non-sampling error."

Accuracy should be a concern when the data-measurement instruments are designed. The evaluation contractor should always pretest questionnaires and other data-collection instruments before deploying them for actual evaluation measurements.

OMB Clearance to Collect Data

If the audience from which you need to collect data does not consist exclusively of Federal Government employees, and the evaluation needs primary data from more than nine members of this audience, including potential customers, then the data collection activity will require clearance from the OMB. Federal Government employees are excluded from the OMB clearance requirement only if the questions to be asked of them involve activities associated with their employment; otherwise, surveys of Federal employees, e.g., as civilian participants in a program, also require OMB clearance.

The legal requirement for OMB clearance of a survey is the Paperwork Reduction Act of 1980 (amended 1995). OMB requires that its clearance review take no less than 30 days and no more than 60 days; however, the entire process can take five to 12 months. An expedited process exists for customer-satisfaction research.

The time required to obtain OMB clearance varies:

- **For customer satisfaction surveys and pretests of other survey instruments,** DOE has an expedited process that, in most cases, takes two to four weeks.[33] The Forms Clearance staff of EIA's Statistics and Methods Group can assist EERE staff with this process.
- **For surveys other than customer satisfaction surveys,** the OMB clearance process takes longer. Currently, the entire clearance process may require five to eight months. EERE clearance applications are submitted to the Records Management Office (IM-11) of DOE's Chief Information Officer.

An OMB clearance is valid for three years.[34] Appendix 5 contains additional information about how to obtain an OMB clearance for data-collection.

Quality Assurance (QA)

Every evaluation study should have a quality assurance process built into it. The process is the responsibility of both the program implementer and the evaluator. For the evaluator, the QA plan should include the steps he or she will take to check the accuracy of the data. These may consist of:

[33] OMB approved generic clearance DOE-887.
[34] Clearances resulting from emergency reviews last six months; however, emergency reviews are unlikely to apply for general evaluations.

- **Checks that certain answers to a survey are internally consistent,** e.g., if a residence does not have access to natural gas, but the respondent says the residence uses natural gas to heat water, the responses are not consistent and should be checked or discarded
- **Pre-tests of survey questionnaires to verify that respondents understand the question,** that respondents are interpreting questions the way you want them to, and that skip patterns are correct
- **A checks of each survey question to ensure that it asks only a single question**
- **Specification of normal and expected ranges for quantitative responses** to measurements so that outliers can be detected
- **Specification of procedures for imputing missing data** within a questionnaire if missing-data imputation is proposed
- **Procedures for estimating** whether non-response to a survey will affect the representativeness of the results
- **Double key-entry of manually collected data** if they will be keyed into an electronic database.

See Step 8 for a further discussion of quality assurance requirements that are relevant to evaluation planning.

Develop a Data Collection Plan

The evaluation should have a plan for data collection that specifies the following:

- **What is the population from which data will be collected** to answer the evaluation questions
- **Which indicators will be quantified** to answer the evaluation questions
- **What kind of data-collection method(s) will be used** to make the measurements (may depend on the proposed method of analysis)
- **Whether a sample or census of the population will be used**
- **If a sample will be used, the target confidence interval**
- **Whether OMB clearance will be required,** and if so, an outline of the procedures for doing so
- **A data quality assurance plan** that provides checks on the reliability and accuracy of the measurements
- **Identification of data that may have value for a future evaluation** and provision for archiving it
- **The schedule for the data-collection task.**

The Data Collection Plan will become part of the Evaluation Plan developed in Step 9.

Step 6c: Select the Analytical Methods for Answering the Evaluation Questions

Step 6 provides:

- **A brief overview of the analytical methods that are available for developing answers to the specific evaluation questions from the data that have been collected**
- **Characterization of the methods by their relative defensibility and cost.**

You can expect to gain from this step guidance that will help you determine whether the methods proposed by an evaluation contractor will meet your needs.

Many analytic methods are available for developing findings from data. Table 10 provides a brief overview of some of the more common analytic methods used to develop evaluation results. If the data were collected by a probability sample, select analysis methods that can make use of the statistical properties of the sample. These methods are identified in the "Typical Use in Evaluation" column of Table 10.

Many of the methods described in Table 10 can be used for more than one type of evaluation. The types of evaluation in which the methods are more commonly used are indicated in the "Comment" column. More than one of the methods listed may be used in the same evaluation analysis. For example, engineering analysis is sometimes used to create an estimate of the energy saved from installing a particular energy conservation measure. The engineering estimate is then used as a variable in a regression analysis to estimate a regression coefficient that will adjust the engineering estimate to reflect the actual savings observed by, e.g., metering.

Table 10.
Principal Analysis Methods Used for General Program Evaluations

Analytical Method	Typical Use in Evaluation	Defensi-bility	Relative Cost	Comment
Case study	Describe the causal chain leading to an outcome.	Low to medium	Low	This is an option if the budget is tightly constrained; however, the ability to deduce defensible findings is usually weak. Typically used in process, outcome, and impact evaluations. Used for R&D program evaluations and deployment program success stories. The latter, in particular, may be valuable for attracting additional participants.
Content analysis	Identify themes that exist in unstructured data, e.g., identify the most frequently sought information from inquiries to a call center, or find the themes in focus group transcripts.	Medium	Low to high	The cost of a content analysis will depend on the number of concepts found that are relevant to the evaluation objectives and the number of data sources that have to be content-analyzed. If the number of sources is large, computer algorithms exist that will help to manage costs. Typically used in process evaluations.

(continued)

Analytical Method	Typical Use in Evaluation	Defensi-bility	Relative Cost	Comment
Meta evaluation: evaluation synthesis	Synthesize the findings from evaluations of similar programs that operated under similar circumstances and use them as findings for the program being evaluated. The synthesized findings may also be used as a benchmark for the program being evaluated.	Low	Medium	Meta evaluations can be labor intensive. It may be costly to search for, assess the relevance of, and extract the relevant findings of other evaluations. The programs whose evaluations are reviewed must be similar to the program under evaluation and their evaluation findings must be relevant to the current evaluation's objectives. Typically used in process and impact evaluations.
Expert judgment Delphi analysis Peer review	These forms of expert judgment can be applied in circumstances where (1) collecting quantitative data might be very difficult or costly, and (2) experts exist who are willing to support the evaluation.	Low to high (wide range)	Low	Delphi analysis is a systematic collection, comparison, and synthesis of judgments from several experts on a subject, e.g., the amount of an outcome that is attributable to the program. If the experts cannot reach agreement on a finding, however, the process may be severely discredited.[35] Typically used in process and impact evaluations.
Cost-benefit analysis	Link program achievements to resources expended.	Low to high	Low to high	Usually, cost-benefit analyses are quantitative. At a high level of program aggregation the evaluation cost is low and its strength is good because quantitative cost data are usually available and direct benefits can be estimated with less effort. But if the analysis is for disaggregated activities that are parts of an overall program strategy, it may be so difficult to disaggregate the costs and benefits to the activity level that the results are open to challenge. The benefits of indirect effects, e.g., gaseous emissions reductions and national security, may be difficult to quantify credibly. Used for cost-benefit evaluations.

(continued)

[35] Delphi analysis must be used in the context of the Federal Advisory Committee Act's (5 USC App 2) restrictions on directing Government-formed advisory committees to reach a consensus, as described in EERE's "Peer Review Guide," July 2004, p. 23. These restrictions do not apply, however, if an evaluation contractor establishes the panel. In many cases, the members of a Delphi panel work independently and are not necessarily directed to produce agreement on a finding.

Analytical Method	Typical Use in Evaluation	Defensi-bility	Relative Cost	Comment
Engineering estimation	Calculate estimates of energy savings or emissions reductions based on engineering, physical, and chemical theory.	Medium	Low to medium	Usually calculated as an average for a set of circumstances encompassing those encountered by the program, then stipulated for all similar circumstances. Energy savings coefficients are often developed by engineering estimation. Typically used in outcome evaluations.
Tabulation & cross-tabulation	Count activities, etc., and place them in categories of interest.	Medium-to-high. Depends on use.	Low	Tabulations are used to report the number of outputs, outcomes, etc., observed. Cross-tabulations report the number of outputs, etc., that occur jointly in two or more categories of interest. Typically used in process, outcome, and impact evaluations.
Correlation	Statistically estimate the strength of a relationship between two indicators.	High	Low	Used to determine the degree of relationship (covariance) between selected output and outcome indicators or any two variables. Typically used in process, outcome, and impact evaluations.
Regression, including econometric and discrete choice analysis	Statistically estimate an equation that calculates the value of an *outcome indicator*, e.g., energy saved, given the value(s) of one or more *output*, *activity*, or *external-factor indicator(s)* used as an independent variable(s), e.g., receipt of a training class, installation of an energy-efficiency measure, energy price.	High	High	A regression equation that includes variables for the known influences on energy usage can estimate the amount of energy saved by program participation (one of the influences). The significance of the regression coefficient for the variable representing participation indicates whether the resulting estimated value of the savings per unit of participation is statistically significant and, therefore, defensible. If data for non-participants are included in the regression analysis, the coefficient of the participation variable can be interpreted as net energy savings. Many variants exist for this method. An evaluation expert proposing to use this method for an evaluation should provide evidence of expertise in its use. For maximum defensibility, an independent QA review of the expert's evaluation plan is advised. Typically used in impact evaluations.

(continued)

Analytical Method	Typical Use in Evaluation	Defensi-bility	Relative Cost	Comment
Differences of means and proportions	Comparison of two or more groups on an indicator of interest.	High	Low	Can be used to compare two groups, e.g., a participant and non-participant group, on how their behavior changed on an indicator of interest during the program period. The analysis should include a test of statistical significance. Typically used in process and impact evaluations.
Survival Analysis	A statistical modeling method used to evaluate persistence of energy savings.	Medium	High	Several models of survival analysis exist. A qualified statistician or econometrician is usually required to choose among them. Typically used in outcome and impact evaluations.

Step 7: Identify the Information That Will Go into the Evaluation Report

Before the evaluation begins, specify the types of reporting outputs that the evaluation must provide. If high-level decision-makers will read the report, it may also be useful to specify the expected outline.

- Consider topics or themes related to the evaluation that would be of interest to the audience receiving the report.
- If different decision-makers need different information from the evaluation, consider writing multiple reports.

NOTE:

It is important to specify the types of information that must be in the report of the evaluation before the evaluation begins.

Appendix 6 contains examples of EERE evaluation report outlines. In general, a report outline includes sections for the following:

- Answers to all of the questions specified for the evaluation.
- Recommended improvements to the program, if relevant (indicate which are high priority compared to others).
- A description of the research design, assumptions, how the data were collected, and the analysis methods. These descriptions should be brief in the main report, where the focus is on the answers to the evaluation questions and recommendations. Put the comprehensive explanations in an appendix.
- Recommended improvements to the evaluation process including limitations of the analysis and any lessons learned about data collection and analysis methods that might aid future evaluations. These can be based on the evaluation contractor's experience and observations during the evaluation process.

Step 8: Establish the Quality-Assurance (QA) Review Process

This step is essential to ensure that the evaluation is defensible with consideration given to the resources that were available for it. It specifies how the data collection, analysis, and reporting activities will themselves be evaluated.

NOTE:

A well-defined quality review process must be in place before the evaluation is begun.

For the EERE manager sponsoring[36] an evaluation study, the following guidance applies to general program evaluation studies:

- Use evaluation contractors who are objective, independent parties with no real or perceived conflict of interest (COI). Contractors who have a long-standing relationship with an EERE program *that includes involvement in daily or routine program implementation and analysis activities* that cannot be divorced from the evaluation activity would not be considered independent without special exception. Contractors should be asked to sign a COI form.
- Evaluation contractors are expected to:
  - Prepare a detailed Evaluation Plan (Step 10).
  - Participate in a review of the draft Evaluation Plan and draft evaluation report.
- For evaluations of large-budget programs and programs attracting special stakeholder interest a panel of three to five independent outside evaluation peers who are not part of the evaluation contractor team and who have no real or perceived COI should be assembled to fully scrutinize the contractor's Evaluation Plan, methodology, and reporting.
- The QA procedure to be used in the evaluation study should be mentioned in the study's statement of work, and in the Evaluation Plan. See Step 10.

Two examples of QA review processes are:

- Establish a ***standing*** peer review panel for the evaluation comprised of 3-5 independent outside experts who provide written review of the Draft Evaluation Plan and participate in a Q&A meeting with the evaluation contractor(s) before the evaluation is conducted. The panel reconvenes to review the Draft Evaluation Report.
- Identify an ***ad hoc*** panel of external evaluation experts to review and provide written comments only on the Draft Evaluation Plan and Draft Evaluation Report.

Example

EERE's Industrial Best Practices Program uses a standing peer review panel for its general program evaluations.

[36] "Sponsoring" means the EERE program provides the funds for a study and has a staff that has responsibility for managing the contract of an independent outside evaluation professional. The evaluation professional conducts the study. It is not an option for general program evaluation studies to be conducted only internally by EERE staff.

As Section 2.2 of this Guide pointed out, judgments of the strength of the evaluation are largely subjective; they depend on the reviewer's own training and experience. The objectivity of the process can be aided by creating a list of specific "criteria" that the reviewers must address. The following list includes criteria that have been proposed for peer reviews by other organizations:[37]

Research Design

- The research questions are well formulated and relevant to the objectives of the evaluation.
- The indicators are credible as measures of the outputs and outcomes being evaluated.
- The research design has validity.
- For statistical methods, the degree of relationship between indicators, tests of significance, and confidence intervals (statistical precision) for sample estimates, were built into the analysis and applied wherever possible.
- The research demonstrates understanding of previous related studies.
- The data collection and analysis methods are credible.

Data Collection

- The data and assumptions about the research design are sound.
- All planned data were collected, or if some values are missing, how they were treated.
- If missing data values were inferred, the inference method was appropriate.
- If a survey was conducted, non-response is accounted for.
- The data collection methods were actually implemented as planned, or if revisions were required by circumstances, they were appropriate and the reasons for the revisions are documented.
- Collected data are provided and their layout documented.

Analysis

- The analysis methods were actually implemented as planned, or if revisions were required by circumstances, they were appropriate and the reasons for the revisions are documented.
- The documentation of the methodology is accurate, understandable, and reasonable.

[37] Many of the standards on this list are taken from RAND Corporation, "Reviewer Guidelines."

Reporting

- The report outline draft is appropriate and likely to present the study findings and recommendations well, and to provide documentation of methods used.
- The draft findings and recommendations in the evaluation report follow logically from the research results and are explained thoroughly.
- The report presents answers to all of the questions asked.

The QA procedures should be included in the Evaluation Plan developed under Step 10. Appendix 7 provides more detailed guidance for establishing QA procedures for general program evaluations.

3.4 Selecting an Evaluation Contractor

Steps 9 through 11 are concerned with selecting an evaluation contractor to perform the general program evaluation and monitoring its performance.

NOTE:

One of the PART questions asks whether an independent evaluation expert has performed a program evaluation.

Step 9: Develop the Statement of Work and Select an Evaluation Contractor

This step summarizes the topics that a good statement of work for an evaluation contractor will cover.

Program managers have a choice on the amount of scope detail that is provided in the statement of work (SOW) for an evaluation contractor. A SOW describes in detail the evaluation requirements. When used in combination with the request-for-proposal (RFP) process, the SOW will produce competition between evaluation suppliers on the research design and price of the evaluation.

A typical SOW outline is shown in Appendix 8. The following are some of the details typically found in an SOW:

- **The objectives of the evaluation**
- **The evaluation questions and their priorities.** If the evaluation will be an outcome or impact evaluation, these questions should relate to the types of direct and indirect outcomes to be evaluated (based on program theory and the logic model, plus discoveries during the evaluation).
- **A requirement that the contractor develop an Evaluation Plan (Step 10) and Quality Assurance Process (Step 8), and a statement that they will be reviewed by outside experts in the field, as appropriate.** The Evaluation and QA Plans can be developed either as part of the contractor's proposal or after contract award. Evaluation and QA Plans developed as part of a proposal aid in

evaluating the scope proposed by the contractor; however, they almost always must be updated after contract award. Therefore, post-award Evaluation and QA Plans should be required even if the proposal includes them. The Evaluation Plans must describe a task structure into which the evaluation research activities will be logically organized for performance and monitoring purposes. The QA Plans must cover data collection, analysis, and report writing. Stipulate that the contractors' bids include resources for these plans.

- **Stipulate that lessons learned from previous EERE and other program evaluations must be incorporated into the Evaluation Plan (see Appendix 9).** This includes expected interactions with other evaluation projects, if any.
- **Degree of initiative asked of bidders with respect to proposing defensible methodologies.**
- **Issues and proposed resolutions for potential problems that may be encountered.** Illustrations of potential problems include, collecting data from states, treatment of attribution (for impact evaluations), designing a probability sample, use of savings ratios, and dealing with potential survey non-response issues. See Appendix 9 on Lessons Learned for Improving the Quality of EERE Evaluation Studies.
- **Reports and other deliverables required**, including periodic performance and budget reporting for the evaluation process. One of the deliverables must be the Evaluation Plan (Step 10). If the Evaluation Plan and draft evaluation report will be subjected to a peer review, state this so that the proposers' bids can take this review into account.
- **Consistency in the use of terminology and between requirements.** If the RFP will use technical terms that a bidder may misinterpret, a glossary will help to reduce the number of follow-on questions.

Program managers sometimes ask bidders to provide examples of evaluation reports to help them assess the ability of the bidders' organization to write clear reports. This may reduce the number of bidders, however, as such reports are often proprietary.

- **Resources that the program manager will provide to the evaluation contractor.** Examples include, participant lists, records of outputs and outcomes, expenditure records, and access to program staff for interviews. Having such resources available can improve the quality and reduce the cost of an evaluation.
- **Evaluation schedule and milestones.** Include a kickoff meeting with the contractor to discuss the above topics.

Other requirements and information may be included if the program manager wants to specify greater detail about the evaluation's requirements:

- **Types of information required when answering individual specific questions**, e.g., counts, averages, proportions
- **Required level of statistical precision for survey results**
- **Required tests of significance for statistical relationships**

- **Data-collection and analysis methodologies** that you expect the contractor to use to answer specific evaluation questions
- **Relevant guidance or references that will give the evaluation expert information about the requirements of Federal program evaluations**, e.g., if the evaluation will be used to comply with PART requirements, the expert should be familiar with PART guidance documents. In particular, the expert should review the document, "What Constitutes Strong Evidence of a Program's Effectiveness?" (www.whitehouse.gov/omb/part/2004_program_eval.pdf).

Sometimes contractor support is needed after the final report is accepted, e.g., to present results to stakeholders. Program managers may also ask the evaluation bidders to propose time and materials rates to provide support related to the evaluation after the project is over. However, such support should never involve correcting errors in the evaluation. That is the contractor's responsibility. (A good peer review should enable you to avoid the need for such corrections.)

Apply the contractor evaluation criteria and recommend a contractor for award. Follow DOE's procurement regulations. At the same time or shortly after the contractor is selected, it is advisable to select an external quality-review team to serve the QA function during the evaluation period, as described in Step 8.

4.0 MANAGEMENT DURING THE EVALUATION

This section of the Guide provides guidance for:

- **Developing an Evaluation Plan**
- **Implementing the evaluation.**

Step 10: Develop an Evaluation Plan and Conduct the Evaluation Study

The Evaluation Plan

The statement of work for the evaluation should specify that the evaluation contractor develop and deliver a written Evaluation Plan. This plan will be used to manage the evaluation and will usually be an expansion of the contractor's proposal, modified as agreed between the program manager and the contractor. The contractor will develop it as soon as the project is awarded, and the plan should be approved before the contractor begins to implement other study activities (unless the program manager approves exceptions so that the contractor can begin collecting time-sensitive data).

NOTE:

It is important to establish a work plan for the evaluation project before active data collection begins. This provides an important tool for managing the evaluation.

The Evaluation Plan should include:

- The logic model from **Step 4**
- The questions stipulated by **Step 5**
- The research design, data-collection plan, and analysis plan that were developed in **Step 6**
- The QA procedures developed in **Step 8**
- The deliverables that will be produced during the project
- The basic content of the evaluation report
- The schedule for performing **Steps 6 through 10**.

The program manager will approve the document, or set of documents, that constitutes the Evaluation Plan and have it disseminated to all who will implement and review the evaluation. It is advisable to have the plan reviewed by outside experts (see Quality-Assurance Review Process, Step 8). The contractor should not begin to implement the evaluation study until the program manager gives final approval to the plan.

Four appendices provide additional information relevant to the development of Evaluation Plans. Appendix 10 contains a model Evaluation Plan outline. Appendix 9 describes lessons learned

for improving the quality of evaluation studies. Appendix 6 illustrates an outline of a typical evaluation report. Appendix 11 contains the American Evaluation Association's guidance on ethical principles for evaluators. These principles should be reflected in the Evaluation Plan's content.

The program manager may find it a useful practice to review this Guide with the contractor prior to the latter's developing the Evaluation Plan. Such a review will promote informed communication between the program manager and contractor and help to assure that the evaluation meets the program manager's needs and EERE's expectations.

Conduct the Evaluation

After the program manager approves the Evaluation Plan the evaluation contractor will begin to implement the evaluation study. Most of the effort to implement the evaluation falls on the contractor's staff; however, there are several tasks the program manager may have to undertake during the evaluation to ensure that the study remains on schedule and achieves its objectives. One of these is monitoring the evaluation contractor's work. The next step describes the program manager's responsibility for monitoring the evaluation contractor's work during the evaluation. Other management tasks include providing support to the evaluation contractor that will enable it to perform one or more of the evaluation tasks. Such support may include:

- Supplying program participant lists
- Supplying copies of program tracking systems
- Preparing letters of introduction to participants from whom the contractor will request data
- Copies of written program procedures
- Copies of prior or related general program evaluations
- Assistance in obtaining OMB clearance for a survey, if applicable
- Access to interview program staff
- Timely review of delivered documents and reports. The evaluation report will become the basis of future action; therefore, it is advisable that the program manager allows time to give careful attention to it.

Step 11: Monitor the Evaluation Contractor During the Study

Everyone knows Murphy's Law: "If anything can go wrong, it will."

This law applies to evaluation projects as completely as it applies to any other management endeavor. It will be as important to monitor the evaluation contractor's work as it is to monitor the implementation of the program itself.

NOTE:

The most carefully designed evaluation can fail to provide defensible information if it is not monitored.

These are some features of an evaluation project that should be monitored to ensure the evaluation delivers useful and defensible results:

- Monitor the evaluation team's performance:
  - Require and hold periodic progress-review meetings. Establish and keep a regular schedule for these meetings. They help the contractor as much as they help the program manager to ensure that the project activities are being performed correctly.
  - Written monthly progress reports are useful monitoring tools.
- Monitor the timeliness of the contractor's achievement of milestones.
- Monitor contractor invoices relative to work progress. The rate of expenditure on the project can provide an early warning sign of problems to come.
- Review all milestone products.
- Meet all of the Government's milestones for deliverables or support that have been promised to the contractor. Because program staff tend to give evaluations lower priority than program operations, the contractor's progress can be delayed by inattention to promised assistance.
- Verify that the elements of the QA Plan (Step 8) are followed.
- If something is not as you expected, conduct a special progress review meeting to confront the issue and develop a resolution before it becomes a serious problem.

- Verify that the draft evaluation report reasonably satisfies the agreed outline.

  - The major findings must answer the priority questions selected for data collection and analysis in Step 5.

  - The method descriptions should be clearly presented in a manner that is understandable by peer-reviewers trained in the research methodologies.

5.0 DISSEMINATING AND APPLYING THE RESULTS

After the contractor submits the evaluation report, there are steps that help to ensure the evaluation is used and its purpose is satisfied. These are:

- **Distribute the results.**
- **Use the results to make decisions about the program.**
- **If results can be leveraged by other program managers, or used in future evaluations, make them readily available.**

General program evaluations are intended to leverage action, not to take up shelf space. They have a purpose.

Step 12: Determine Distribution of Final Report and Results and Distribute

Send copies of the evaluation report(s) or notices of their availability to all of those who can use the results. This audience is not limited to the decision-makers who are waiting for the results to inform their decisions. It could also include stakeholders in the program and other program managers who might benefit from the findings.

NOTE:
Ensure that those who need to use, or can use, the evaluation findings receive them.

Develop an action plan for disseminating the evaluation findings:

- Make presentations as needed to decision makers, program staff, implementers, and stakeholders. Be proactive about getting the results noticed and utilized. The publicity given to the findings may help the program.
- Share lessons learned about the evaluation process with other program managers.

Step 13: Make, or Monitor the Making of, Decisions about the Program Based on the Evaluation Results

A variety of applications for the evaluation results exist:

Now that you've finished the evaluation, use the results to benefit your program.

- First, and foremost, use the results to make decisions to improve the program.
- If the results will inform decisions made by others, the results give you the opportunity to also supply them with plans for improving the program, if appropriate.
- One of the PART Section IV (Program Results) questions asks, "Does the program demonstrate improved efficiencies and cost effectiveness in achieving program goals

each year?" Evaluation results can supply invaluable information that will help you answer this question in the affirmative.

- One of the most common lessons learned from past EERE evaluations is the importance of establishing adequate program record keeping. Good program record keeping helps management monitor program progress and helps to reduce the cost of future general program evaluations. Use the experience in conducting the evaluation to improve program record keeping, e.g., customer contact information, outputs produced by market segment.

Table 11 gives some examples of actions taken by three EERE programs that were based on general program evaluation findings that have resulted in benefits to the programs.

Table 11.
Illustrations of EERE Program Improvements Resulting from General Program Evaluations

Sample Evaluation Recommendation	Program Response/Action	Program Benefit
Industrial Assessment Center (IAC) Program Evaluation (1999)[38]		
1. Improve the maintenance of client-contact information	a. Record contact name and title when the energy assessment is scheduled and update it when the assessment-recommendation (AR) implementation report is uploaded. b. Local IACs maintain a client mailing list.	a. The AR-implementation database reduces the effort and time needed by headquarters and IAC staff to quantify and report the benefits of program activities and outputs. b. The mailing list improves the effectiveness of the IAC Program at the local level by making it possible to keep interested clients informed of new energy-efficiency technologies and program services. c. Existence of contact information reduces the cost of future evaluations by reducing the effort needed to identify who can provide information on actions taken by the client.
2. Maintain an updated database as new information is received about replication, spin-off, and AR status	a. Record client information on whether ARs were implemented and whether additional efforts, such as replication and spin-off, were implemented.	a. Information on replication and spin-off identifies long-term outcomes (technology diffusion beyond the original assessment) thereby increasing the energy savings benefits that may be reported by the IAC Program. b. Recording this information soon after implementation helps to ensure that the information is accurate. This improves the defensibility and reduces the cost of future general program evaluations. c. The centers use this information to improve their understanding of how clients (1) implement larger efficiency projects, and (2) replicate and spin-off measures after the assessment.

[38] Oak Ridge National Laboratory, "Industrial Assessment Center Program Impact Evaluation," ORNL/CON-473, December 1999.

Sample Evaluation Recommendation	Program Response/Action	Program Benefit
Industrial Best Practices Program Evaluation (2002)[39]		
1. Keep a record of attendees at training sessions.	a. Maintain a database of all trainees.	a. The database of trainees improves the effectiveness of the program by making it possible to inform interested trainee organizations of new energy-efficiency technologies and program services.
2 Keep records of the number of materials downloaded from the Web site and of the participants who register to use the program's software.	a. This information is now posted to the Best Practices Tracking System.	a. The existence of this information in a database reduces the effort and time needed by headquarters and IAC staff to manage and report on program outputs. b. All of the information described above reduces the cost of future evaluations by reducing the effort needed to identify who may have benefited from the Program's outputs.
FEMP Evaluation (2001)[40]		
1. Market FEMP's services more actively to building operations and maintenance (O&M) personnel because these have been less involved in the program in the past. They might be able to increase individual agencies' commitments to improved efficiency at relatively low incremental cost.	a. Created an O&M subprogram that has produced a number of educational documents and workshops. The documents include a "Continuous Commissioning Guide," a "Best Practices Guide for Efficiency Operations and Maintenance," a series of "O&M First" fact sheets to address low-cost/no-cost ways to save on heating, cooling and water, and more. b. Created a full workshop track for O&M at its annual workshop and exposition. c. Developed an O&M Web site. d. Developed a manual on contracting for a Resource Efficiency Manager and an associated Web-based training program.	a. The evaluation recommendation led FEMP to identify and address the needs of a Federal market that had heretofore been underserved. This has increased the effectiveness of the program by helping agencies make progress toward the Federal government's energy-usage reduction goals at relatively low cost using existing staff. b. Agencies have showed strong interest in taking advantage of the new educational activities. The full O&M track at the annual workshop and exposition had record attendance in 2005 and surpassed all other tracks in total attendance. The Best Practices Guide is the most popular download from the FEMP Web site.

Develop and implement a plan to assume ownership of the evaluation's results. Use the plan to improve program operations, inform decision-making, improve record keeping, and track progress.

[39] Oak Ridge National Laboratory, "Preliminary Estimation of Energy Management Metrics for the Best Practices Program," ORNL/TM-2002/134, July 2002.
[40] TecMRKT Works and Sandia National Laboratories, "2001 FEMP Customer Survey," June 2001.

Step 14: Establish/Update Program Records for Use in Future Evaluations

Each general program evaluation is an investment of program resources to formally assess program achievements and identify recommendations for program improvements. The evaluation experience itself is valuable because it shortens the learning curve for the next evaluation. To preserve this knowledge and associated learning for future use, the program manager should:

Develop a program tracking system for activities, outputs, and, if the information is available, outcomes to improve the effectiveness of day-to-day management and facilitate future evaluations.

- **Develop and maintain a program-tracking database** containing routinely collected participant information for use in future evaluations.
- **Establish a database** for the achievements and recommendations.
- **Archive discussions of the problems encountered** in the evaluation and how they were resolved.
- **Archive the updated contact information** on the individuals and organizations contacted.

Examples

Six past evaluation studies contained a recommendation to develop a program-tracking database to routinely collect the types of data needed for improved day-to-day program coordination and management, and for future evaluation activities. These evaluations were conducted for the Industrial Assessment Center (IAC) Program, the Best Practices (BP) Program, Building America, Rebuild America, FEMP, and the former Motor Challenge Program). Here are three examples:

- **Motor Challenge Program Evaluation:**
  "Program record keeping must be enhanced to enable managers and implementation staff to better characterize establishments quickly as to their function (end-user, vendor, utility, or trade association), industry, and size. This will aid in program marketing, client relations management, and evaluation."
- **IAC Evaluation:**
  "Improved maintenance of client contact info, update database as new information is gathered from long-term follow-up, on replication, spin off and AR status."
- **Best Practices Program Evaluation:**
  "Submit training attendee lists, track materials acquired via the Web, identify training sessions that also include [estimates of potential energy savings in specific facilities], assist with characterizing BP clients, and maintain consistent records."

6.0 SUMMARY

The emphasis on good management practice and on the demonstration of program results in return for public resources creates a need for good general program evaluations. This Guide provides the program manager with the basics for understanding the general program evaluation process and how to manage it. Current OMB requirements specify that independent evaluation experts perform general program evaluations, but after the evaluation results have been reported, the ball is back in the program manager's court. Besides satisfying OMB's requirements, a good evaluation will produce information that can help the manager operate a better, more efficient program. It is up to the manager to make use of it.

APPENDICES
EERE Guide for Managing General Program Evaluation Studies

Getting the Information You Need

U.S. Department of Energy
Energy Efficiency and Renewable Energy
Bringing you a prosperous future where energy is clean, abundant, reliable, and affordable

Appendices

Appendix 1
Checklist for Managing a General Program Evaluation Study

Checklist for Managing a General Program Evaluation Study[1]

Status	Action	Step Reference(s)
Initial Planning		
	Determine program elements subject to evaluation	Step 1
	Determine the types of decisions to be informed by the evaluation	Step 1
	Develop statement(s) of evaluation objectives	Step 1
	Construct or confirm program logic model	Step 4
	Determine type(s) of information needed to evaluate objectives	Step 1
	Determine appropriate type(s) of evaluation	Step 1
	Determine resources required for the evaluation	Step 2
	Identify available resources (budget, staff, schedule constraints) for conducting the evaluations	Step 2
	Reconcile resources and requirements	Step 2
	Determine date that final evaluation results will be needed to contribute to decisions	Step 3
	Develop procurement/award/implementation schedule to yield evaluation results by decision date	Step 3
Evaluation Design and Procurement		
	Develop list of general questions	Step 5
	Develop list of specific researchable questions based on the general questions	Step 5
	Develop evaluation report outline	Step 7
	Develop Statement of Work (SOW), including: Objectives Research design requirements Deliverables Quality assurance requirements Schedule/milestones Evaluation Plan requirements	Step 9 Step 1 Step 6 Steps 7, 9 Steps 6, 8 Step 9 Steps 9, 10
	If survey(s) requested, determine OMB clearance requirements Full clearance Customer satisfaction generic clearance	Step 6b
	Follow other appropriate DOE contractor procurement procedures for hiring an evaluation contractor	Step 9

[1] This checklist has been developed from, and all references are to, the *EERE Guide for Managing General Program Evaluation Studies* (February 2006). The Guide notes that several of the steps should be taken concurrently or revisited after later steps are performed. The sequence of the steps in this checklist reflects this guidance.

Status	Action	Step Reference(s)
		(Continued)
	Establish a quality assurance (QA) review process, including criteria	Step 8
	Review proposals and select evaluation contractor	Step 9
Implementation of Study		
	Set progress review meeting schedule	Step 9
	Set progress reporting schedule	Step 9
	Review Evaluation Plan including: Internal review Outside expert review (if any)	Step 10
	Approve Evaluation Plan	Step 10
	Authorize evaluation contractor to implement Evaluation Plan	Step 10
	Monitor the evaluation contractor's work	Step 11
	Review and approve reports	Step 11
Using Evaluation Results		
	Develop distribution list for final report (ensure that all stakeholders receive a copy of the evaluation findings that are of interest to them)	Step 12
	Develop action plan to disseminate evaluation findings	Step 12
	Utilize evaluation results in program decisions	Step 13
	Establish/update program records for use in future evaluations	Step 14

Appendix 2
Evaluation-Relevant Sections of OMB's PART Instructions

Evaluation-Relevant Sections of OMB's PART Instructions

Introduction

The Office of Management and Budget (OMB) has established the Program Assessment Rating Tool (PART) as a formal link between budget decisions and program performance. The PART is intended to evaluate a program's purpose and design, planning, management, and results and to determine its overall effectiveness and accountability. General program evaluations constitute a major input for PART.

The rating-tool feature of PART asks a series of questions designed to provide a consistent approach to rating programs across the Federal government. They are designed to assess whether a program has sound management practices and is producing results. The answers to these questions rely on objective data to assess programs on a range of issues related to performance.

This appendix excerpts the PART questions for which general program evaluations will provide, and are sometimes *required* to provide, answers. It paraphrases the PART instructions on how to answer them. These instructions are found in "*Instructions for the Program Assessment Rating Tool, PART Guidance for FY2006 Budget*" (*PART Instructions*). *PART Instructions* is issued by the Office of Management and Budget and can be downloaded from www.whitehouse.gov/omb/part/2006_part_guidance.pdf.

The purpose of this appendix is to provide more information on the role that general program evaluations play in the PART requirements than could be included in the *EERE Guide for Managing General Program Evaluation Studies (Guide)*. This appendix is not intended to replace *PART Instructions*. Program managers should still review the appropriate sections of *PART Instructions* that are identified in this appendix.

The questions that constitute the PART are generally written so that they can be answered in a *Yes/No* format. They require the user to explain the answer briefly and to include relevant supporting evidence. Responses must be evidence-based and may not rely on impressions or generalities. General program evaluations can provide this evidence. Where hard evidence is unavailable, assessments can rely on professional judgment, which can also be developed from general program evaluations.

Eight of the questions included in *PART Instructions* can specifically benefit from information developed from general program evaluations. These have been selected for this appendix. Other questions may also benefit. The selected questions (1) can be answered directly or indirectly using one or more of the general program evaluations described in the *Guide*, and (2) are relevant to EERE deployment and R&D programs. PART identifies questions that are unique to R&D with "RD" in the question number, e.g., "3.RD1."

Two of the PART questions (#2.6 and #4.5) specifically require that a program perform what OMB calls "independent evaluations." Independent evaluations are the same as the "general program evaluations" described in the *Guide*. For the other questions included in this

appendix, general program evaluations can be used to develop evidence to support a favorable (or unfavorable) answer to the question.

The questions in *PART Instructions* are divided into four topic areas: (1) program purpose and design, (2) strategic planning, (3) program management, and (4) program results/accountability. The PART questions in this appendix are divided into four sections corresponding these topics.

The guidance for answering each of the selected questions is provided in subsections under each question that cover:

- The types of answers that are acceptable for the questions in each topic area
- The types of evidence needed to support the acceptable answers, e.g., a *Yes* answer
- Evidence and data required to support the answer
- For some of the questions, as identified in *PART Instructions*, an explanation of some of the terms used in the evidence and data subsection.

These subsections and their information are excerpted directly from (but in some cases paraphrased for brevity) *PART Instructions*. References with page numbers have been added to indicate when information is paraphrased and when it is directly quoted from the *PART Instructions*. These references can also help the program manager find the complete PART instructions for a question. For some of the questions, a text box contains EERE commentary relating the guidance to specific types of general program evaluation.

The structure of the guidance in this appendix for the questions is as follows:

Topic Area

Possible Answers to Questions Asked for the Topic Area (e.g., Yes, No, Not applicable)

Question #n.n

Elements of a *Yes* Answer

Elements of a *No* Answer

***Not Applicable* Answer**

Evidence/Data

Elements of an Acceptable Answer (information on terminology used in some of the questions)

Comment: EERE Commentary

The PART questions included in this appendix are:

1. Question #1.4: Is the program design free of major flaws that would limit the program's effectiveness or efficiency?
2. Question #1.5: Is the program design effectively targeted, so that resources will reach intended beneficiaries and/or otherwise address the program's purpose directly?
3. Question #2.6: Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the program, interest, or need?
4. Question #3.1: Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?
5. Question #3.RD1: For R&D programs other than competitive grants programs, does the program allocate funds and use management processes that maintain program quality?
6. Question #4.1: Has the program demonstrated adequate progress in achieving its long-term performance goals?
7. Question #4.3: Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?
8. Question #4.5: Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

1. Program Purpose and Design

Possible Answers to Questions for This Topic: "Options for answers are *Yes*, *No* or *Not Applicable*. Design flaws in the underlying legislation can and should be considered and supported by evidence, and are grounds for a *No*. *Not Applicable* answers are likely to be rare, … as these questions should apply to virtually all programs." (Direct quote: *PART Instructions*, p. 14)

Question #1.4 Is the program design free of major flaws that would limit the program's effectiveness or efficiency?

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 16)
- No strong evidence that another approach or mechanism would be more efficient or effective to achieve the intended purpose
- Program structure continues to make sense given changing conditions in the field.
- Program impact is extended by leveraging funds and contributions from other parties.

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 16)
- Evidence that another approach or mechanism would be more efficient or effective to achieve the intended purpose.

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 14)
- Not an option except in rare cases.

Evidence/Data
Evidence demonstrating efficient design can include cost effectiveness studies comparing alternative mechanisms with the current design. Evidence on relative benefits and costs of the activity are also useful. (Paraphrase: *PART Instructions*, p. 16)

Comment: The principle concern of Question #1.4 is comparing alternative program designs. Process and cost-benefit general program evaluation studies can provide evidence to support the answer to this question.

Question #1.5 Is the program design effectively targeted, so that resources will reach intended beneficiaries and/or otherwise address the program's purpose directly?

"Unlike Question #1.4, which addresses examination of alternatives to achieve a program's goals, this question asks whether program resources under the *chosen* alternative are oriented toward the effective achievement of the program's purpose." (Direct quote: *PART Instructions*, p. 17)

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 17)
- Demonstrate that the right beneficiaries are being targeted, activities that would have occurred without the program are not subsidized, and program funds are targeted effectively to meet program purposes
- Acceleration of activities due to Federal funding can be grounds for a Yes, but there should be evidence that the acceleration warrants the subsidy or application of funding.

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 17-18)
- Lack of evidence to support a Yes answer or evidence that contradicts the elements of a *Yes* answer
- Programs not designed to avoid unwarranted shares of funding going to beneficiaries who do not need or merit the funding.

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 14)
- Not an option except in rate cases.

Evidence/Data
Evidence should show that the program is designed to (1) reach the highest practicably percentage of target beneficiaries, (2) not subsidize outcomes that would have occurred anyway, and (3) have the smallest practicable share of funds going to unintended beneficiaries. (Paraphrase: *PART Instructions*, p. 17-18)

Comment: All of the types of general program evaluations can provide evidence to support the answer to this question.

2. Strategic Planning

Possible Answers to Questions for This Topic: "Options for answers are *Yes*, *No* or *Not Applicable*. While it is recognized that some programs may have greater difficulty than others in developing quantitative performance goals, programs must have meaningful and appropriate methods for demonstrating results." (Direct quote: *PART Instructions*, p. 18)

Question #2.6 Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 23)
- Regularly scheduled objective, high-quality, independent evaluations
- Evaluations that examine how well the program is accomplishing its mission and meeting its long-term goals
- Program in the process of developing new evaluation approaches that will provide the most rigorous evidence possible by a specified future date (if rigorous evaluation plan not already in operation).

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 26)
- Absence of any independent evaluation, unless there are plans to carry out a rigorous program evaluation in the near future
- Evaluation data exists but is not from an independent source
- Insufficient independent evaluation data
- Evaluations that address process and not performance related to goals.

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 23)
- Not an option as a response to this question. All programs should be capable of undergoing some type of evaluation acceptable to OMB.

Evidence/Data
"Evidence should include a program evaluation plan or schedule of program evaluations and program documentation describing the type of evaluation, including scope and quality, and the criteria for selecting an independent evaluator." (Direct quote: *PART Instructions*, p. 26)

Elements of an Acceptable Evaluation
Quality. "… agencies should provide evidence that they have chosen and applied evaluation methods that provide the most rigorous evidence of a program's effectiveness that is appropriate and feasible." (Direct quote: p. 24) "The most significant aspect of program effectiveness is *impact* – the outcome of the program, which otherwise would not have occurred without the program intervention." (Direct quote; emphasis in original: *PART Instructions*, p.24) "Overall, *evaluations must be appropriate to the type of program*." (Direct quote; emphasis in original: *PART Instructions*, p.24)

Independence. Evaluations are conducted by independent, non-biased parties with no conflict of interest. Evaluations conducted by the program itself are generally not considered

to be independent. However, contracting the evaluation to a third party paid by the program may qualify as independent. (Paraphrase: p. 26) "OMB examiners and agency staff will determine if a specific evaluation can be considered 'independent' for this question." (Direct quote: *PART Instructions*, p. 26)

Scope. "This question looks directly at whether there are evaluations of the program's achievement of performance targets, and these evaluations examine the underlying cause and effect relationship between the program and the target. In cases where a comprehensive evaluation is unnecessary based on the known effectiveness of an intervention and performance data on the program, evaluations that fill gaps in performance information can meet the elements of a *Yes* answer. A program's *effectiveness*, including *impact*, also may be considered. … Evaluations also should include recommendations on how to improve the program's performance. **To ensure the program continues to meet its performance targets, an evaluation should be scheduled on a periodic basis, such as every two to five years, or whatever time schedule is reasonable based on the specific program and its mission and goals**." (Direct quote; italicized emphasis in original, bold emphasis added: *PART Instructions*, p. 25)

3. Program Management

Possible Answers to Questions for This Topic: Options for answers are *Yes*, *No* or *Not Applicable*. (Direct quote: p. 30)

Question #3.1 Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 31)
- Program regularly collects high-quality performance data relating to key program goals and uses that information to adjust program priorities, allocate, resources, or take other appropriate management actions
- Program has collected the baseline performance data necessary to set meaningful, ambitious performance targets.

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 31)
- Lack of evidence that timely and credible performance information is collected and used to improve performance.

***Not Applicable* Answer**
- No specific OMB guidance.

Evidence/Data
"Evidence can include a description of how the agency uses performance information in managing the program, as well as illustrative examples of recent management actions based on performance information. Evidence can also include steps taken by a program to enact necessary improvements cited by a specific evaluation." (Direct quote: *PART Instructions*, p. 31)

Elements of Acceptable Performance Information
"Timely performance information is information that reflects current performance and is current enough to be useful in program management. Credible performance information is information that is collected through a systematic process with quality controls to confirm the validity of the data." (Direct quote: *PART Instructions*, p. 31)

Comment: Needs/market assessment evaluations can provide baseline data against which subsequent performance data can be compared to identify trends. Trends will help managers assess progress and identify whether a need to improve performance exists. The other types of general program evaluation can provide the subsequent credible performance data used for the trends.

Question #3.RD1 For R&D programs other than competitive grants programs, does the program allocate funds and use management processes that maintain program quality?

The purpose of this question is to determine whether the program uses a clearly stated, defensible method for allocating its R&D funding. (Paraphrase: *PART Instructions*, p. 43)

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 43)
- Program allocates funding using a broadly competitive process based on merit, or has compelling justifications for R&D funding allocated through other means
- All program funds allocated through means other than unlimited competition must document the processes they use to distribute funds to each type of R&D performer (e.g. federal laboratories, federally funded R&D centers, universities, etc.).

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 43)
- Absence of the elements of a *Yes* answer.

***Not Applicable* Answer**
- No specific OMB guidance.

Evidence/Data
"Evidence can include a description of the awards process, percentage of funds earmarked, percentage of funds subject to competitive peer review, and results of external assessments." (Direct quote: *PART Instructions*, p. 43)

"Programs are encouraged to use external assessment of the methods they use to allocate R&D and maintain program quality." (Direct quote: *PART Instructions*, p. 43)

Comment: Process evaluations can provide the external assessment evidence mentioned in the instructions for this question.

4. Program Results/Accountability

Possible Answers to Questions for This Topic: Answers in this section are rated as *Yes*, *Large Extent*, *Small Extent*, and *No*. *Not Applicabl*e answers might be an option depending on the specific question. (Paraphrase: *PART Instructions*, p. 44)

Question #4.1 Has the program demonstrated adequate progress in achieving its long-term performance goals?

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 44-45)
- Program is on track to meet all the long-term performance goals – including ambitious targets and timeframes
- Partial credit, such as *Large Extent* or *Small Extent*, should be given in cases where there is partial, but notable, achievement of long-term targets.

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 45)
- Program not making progress toward achieving long-term goals (even if achieving annual targets).

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 44)
- In general, not appropriate as a response to this question.

Evidence/Data
"Evidence can include data from the agency's GPRA performance report, a strategic plan, or other Administration goals and objectives. Reports detailing customer satisfaction with program performance, program reports detailing rates of utilization or participation, or independent evaluations of the program's performance may also be considered as relevant evidence." (Direct quote: *PART Instructions*, p. 45)

Comment: Process, outcome, and impact studies can provide strong evidence that one or more goals have been achieved. See also Question #4.5.

Question #4.3 Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?

Elements of a *Yes* Answer (Paraphrase: *PART Instructions*, p. 46)
- Program demonstrates improved efficiency or cost effectiveness over the prior year. Efficiency improvements should generally be measured in terms of dollars or time
- Programs that clearly demonstrate very high levels of efficiency through other means may receive a *Yes* without documenting increasing efficiency over time.

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 46)
- Program does not demonstrate improved efficiencies or cost effectiveness over the prior year.

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 44)
- This is a potential answer if the program is already operating at very high efficiency levels.

Evidence/Data
"Evidence can include meeting performance targets to reduce per unit costs or time, meeting production and schedule targets; or meeting other targets that result in tangible productivity or efficiency gains." (Direct quote: *PART Instructions*, p. 46)

Comment: Process evaluations and cost-benefit analyses can provide evidence of efficiencies and cost effectiveness.

Question #4.5 Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

Elements of a *Yes* Answer

- An evaluation completed according to the standards set in Question #2.6 that indicates the program is effective. (Paraphrase: *PART Instructions*, p. 47) "The most definitive data supporting a program's overall effectiveness would be from a randomized controlled trial, when appropriate and feasible. Data from other evaluation methods, such as quasi-experimental and non-experimental, can be considered as detailed in Question 2.6, but should be scrutinized given the increased possibility of an erroneous conclusion. If a program is taking necessary steps to correct deficiencies uncovered by the evaluation, the user should address this effort in Question 3.7." (Direct quote: *PART Instructions*, p. 47)

Elements of a *No* Answer (Paraphrase: *PART Instructions*, p. 47)

- Absence of any independent evaluation of program effectiveness, or an evaluation that does not indicate that the program is effective.

***Not Applicable* Answer** (Paraphrase: *PART Instructions*, p. 48)

- Not an option as a response to this question. All programs should be capable of undergoing some type of evaluation of effectiveness that is acceptable to OMB.

Evidence/Data

"Evidence can include findings of an evaluation conducted by the General Accounting Office, Inspectors General, academic and research institutions, agency contracts, and other independent entities." (Direct quote: *PART Instructions*, p. 48)

Elements of an Acceptable Evaluation

"Relevant evaluations would be at the national program level, rather than evaluations of one or more program partners, and would not focus on process indicators such as the number of grants provided, or hits on a web site. Relevant evaluations would consider a program's impact and effectiveness. Evaluations conducted by the program itself should not be considered 'independent.' However, if the program has contracted out the evaluation to a third-party, it might be considered independent. Evaluations conducted by an agency's Inspector General or program-evaluation office also might be considered 'independent.' OMB examiners and agency staff will determine if a specific evaluation can be considered 'independent' for this question." (Direct quote, *PART Instructions*, p. 47-48)

An acceptable evaluation would also meet the quality, scope, and independence criteria included in the guidance for Question #2.6. (Paraphrase, *PART Instructions*, p. 48)

Appendix 3
Glossary of General Program Evaluation Terminology

Glossary of General Program Evaluation Terminology

The definitions in this glossary are adapted from Federal, state, and academic sources such as those listed in the bibliography at the end of this appendix.

Accuracy	The correspondence between the measurement made on an indicator and the actual value of the indicator at the time of measurement.
Activities	A term used generically with logic modeling to describe the action steps necessary to produce program outputs.
Bias	The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value.
Comparison Group	A group of individuals or organizations that have not had the opportunity to receive program benefits which has been selected because its characteristics match those of another group of individuals or organizations that have had the opportunity to receive program benefits. The characteristics used to match the two groups should be associated with the action or behavior that the program is trying to promote. In evaluation practice, a comparison group is often used when random selection of recipients of the program benefit and a control group is not feasible.
Construct	An attribute, usually unobservable, such as attitude or comfort that is represented by an observable measure.
Control Group	A randomly selected group of individuals or organizations that have not had the opportunity to receive program benefits. A control group is measured to determine the extent to which its members have taken actions promoted by the program. These measurements are used to estimate the degree to which the promoted actions would have been taken if the program did not exist.
Cost-Benefit and Cost-Effectiveness Evaluation	Comparison of a program's outputs or outcomes with the costs (resources expended) to produce them. Cost-effectiveness analysis assesses the cost of meeting a single output, objective, or goal, and can be used to identify the least costly alternative to meet that output, objective, or goal. Cost-benefit analysis aims to identify and compare all relevant costs and benefits, usually expressed in dollar terms. The two terms are often interchanged in evaluation discussions.
Cross-Sectional Data	Observations collected on subjects or events at a single point in time.

Deemed Savings	An estimate of an energy savings or energy-demand savings outcome (gross savings) for a single unit of an installed energy-efficiency or renewable-energy measure that (1) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose, and (2) will be applied to situations other than that for which it was developed. That is, the unit savings estimate is "deemed" to be acceptable for other applications. Deemed savings estimates are more often used in program planning than in evaluation. They should not be used for evaluation purposes when a program-specific evaluation can be performed. When a deemed savings estimate is used, it is important to know whether its baseline is an energy-efficiency code or open-market practice. The most extensive database of deemed savings is California's Database for Energy Efficiency Resources (DEER). The deemed savings in DEER are tailored to California. http://eega.cpuc.ca.gov/deer/
Defensibility	The ability of evaluation results to stand up to scientific criticism. Defensibility is based on assessments by experts of the evaluation's validity, reliability, and accuracy. See also Strength.
Direct customers	The individuals or organizations that receive the outputs of a program.
External Validity	The extent to which a finding applies (or can be generalized) to persons, objects, settings, or times other than those that were the subject of study.
Generalizability	Used interchangeably with "external validity."
Impact Evaluation	The application of scientific research methods to estimate how much of the observed results, intended or not, are caused by program activities and how much might have been observed in the absence of the program. This form of evaluation is employed when external factors are known to influence the program's outcomes in order to isolate the program's contribution to achievement of its objectives.
Indicator (also Performance Indicator)	A particular characteristic used to measure outputs or outcomes; a quantifiable expression used to observe and track the status of a process. An indicator constitutes the observable evidence of accomplishments, changes made, or progress achieved.
Internal Validity	The extent to which the causes of an effect are defensibly established by an inquiry.
Logic Model	A plausible and sensible diagram of the sequence of causes (resources, activities, and outputs) that produce the effects (outcomes) sought by the program.

Longitudinal Data	Observations collected over a period of time. The sample (instances or cases) may or may not be the same each time but the population remains constant. Longitudinal data are sometimes called "time series data."
Market Effects	A change in the structure or functioning of a market or the behavior of participants in a market that results from one or more program efforts. Typically the resultant market or behavior change leads to an increase the adoption of energy-efficient or renewable-energy products, services, or practices. Examples include an increase in the proportion of energy-efficient models displayed in an appliance store, the creation of a leak inspection and repair service by a compressed-air-system vendor, an increase in the proportion of commercial new-construction building specifications that require efficient lighting.
Measurement	A procedure for assigning a number to an observed object or event.
Needs/Market Assessment Evaluation	Measurement of those needs of the program's customers that are within, or may be within, the scope of the program's objectives. This form of evaluation examines whether a need exists to change the program's activities and outputs in order to achieve more effective outcomes.
Outcome	A term used generically with logic modeling to describe the effects that the program seeks to produce. It includes the secondary effects that result from the actions of those it has succeeded in influencing.
Outcome Evaluation	Measurement of the extent to which a program achieves its outcome-oriented objectives. It measures outputs and outcomes (including unintended effects) to judge program effectiveness but may also assess program process to understand how outcomes are produced.
Output	A term used generically with logic modeling to describe all of the products, goods, and services offered to the programs direct customers.
Panel Data	A special form of longitudinal data in which observations are collected on the same sample of respondents over a period of time.
Probability Sampling	A method for drawing a sample from a population such that all possible samples have a known and specified probability of being drawn.
Process (or Implementation) Evaluation	Assessment of the extent to which a program is operating as intended. It assesses program activities' conformance to statutory and regulatory requirements, to program design, and to professional standards or customer expectations.

Program Evaluation (also General Program Evaluation)	Program evaluations are independent systematic studies conducted periodically on an ad hoc basis to assess how well a program is working and whether it is achieving its intended objectives. They are conducted by experts external to the program staff.
Qualitative Data	Information expressed in the form of words.
Quantitative Data	Information expressed in the form of numbers. Measurement gives a procedure for assigning numbers to observations. See Measurement.
Random Assignment	A method for assigning subjects to one or more groups by chance.
Reliability	The quality of a measurement process that would produce similar results on: (1) repeated observations of the same condition or event; or (2) multiple observations of the same condition or event by different observers.
Representative Sample	A sample that has approximately the same distribution of characteristics as the population from which it was drawn.
Simple Random Sample	A method for drawing a sample from a population such that all samples of a given size have equal probability of being drawn.
Significance Level	The probability of finding a relationship between two sampled characteristics such as the program treatment and an outcome, when, in fact, there is no relationship.
Strength	A term used to describe the overall defensibility of the evaluation as assessed by use of scientific practice, asking appropriate evaluation questions, documenting assumptions, making accurate measurements, and ruling out competing evidence of causation.
Structured Interview	An interview in which the questions to be asked, their sequence, and the detailed information to be gathered are all predetermined; used where maximum consistency across interviews and interviewees is needed.
Treatment Group	The subjects of the intervention being studied.
Validity	See Internal Validity and External Validity.

Bibliography

1. California Public Utilities Commission, "The California Evaluation Framework," June 2004. www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf

2. DOE," The Performance-Based Management Handbook, Volume 4, November 2000, Appendix A. www.orau.gov/pbm/pbmhandbook/pbmhandbook.html

3. GAO, "Designing Evaluations," GAO/PEMD-10.1.4, March 1991, pp. 92-94. www.gao.gov/special.pubs/pe1014.pdf

4. GAO, "Performance Measurement and Evaluation: Definitions and Relationships," GAO/GGD-98-26, April 1998. www.gao.gov/special.pubs/gg98026.pdf

5. OMB, "Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports, Circular No. A-11, Part 6, June 2002, section 200-2. www.whitehouse.gov/omb/circulars/a11/churrent_year/parta6.pdf

6. OMB, "Instructions for the Program Assessment Rating Tool," pp. 7-10. www.whitehouse.gov/omb/part/2006_part_guidance.pdf

7. McLaughlin, J.A., and Jordan, G. B., "Logic Models: A Tool for Telling Your Program's Performance Story," *Evaluation and Program Planning*, Volume 22, Number 1, February 1999.

8. University of Wisconsin Extension, "Planning a Program Evaluation," February 1996, pp. 2-10. www.uwex.edu/ces/pdande/evaluation/evaldocs.html

Appendix 4
General-to-Specific Evaluation Questions

General-to-Specific Evaluation Questions

Introduction

Once evaluation objectives are established, the research needs to be framed into general and specific questions that can be the specific subjects of the research planning and evaluation effort. General questions are derived from the evaluation objectives. Each general question implies certain specific research questions that represent it. The specific questions are questions that are capable of being answered through data collection and analysis. The following sets of general and specific questions are grouped by type of evaluation:

- Needs/market assessment
- Process, or implementation
- Outcome
- Impact
- Cost-benefit.

These general and specific questions are offered as examples of the kinds of questions addressed by the different types of general program evaluations.

A. Needs/Market Assessment Evaluation

General Question 1: What additional customers and markets could be served?

a. What are the currently unserved populations and market segments that could benefit from the program?

b. Are there additional delivery channels that could be used to reach the target populations?

General Question 2: What do customers need that is not currently being provided?

a. What gaps currently exist in the services available to target populations?

b. What specific tools and services are needed by customers that are not provided by the program?

General Question 3: What is the market baseline?

a. What are the key market segments?

b. Who are the key market actors and how do they interact?

c. What is the current extent of market penetration for the program's targeted technologies?

d. What is the nature and magnitude of current market barriers to the greater use of technologies or practices promoted by the program?

B. Process or Implementation Evaluation

General Question 1: Is program design and organization adequate?

a. Are program goals too high? Too low?

b. What populations and market segments are being served, and through what delivery channels?

c. Is it easy for customers to join or participate in the program?

d. What motivates customers to participate?

e. Are program delivery strategies consistent with customer motivations?

f. Do marketing materials emphasize benefits that have high value for customers?

g. Do the characteristics of the available tools and services allow for their easy adoption?

General Question 2: Is the program producing the outputs it was intended to produce?

a. What is the level of awareness of energy efficiency and renewable energy opportunities in target populations?

b. Are customers participating at expected levels? Are some customer groups participating more than others? Why?

c. Which tools and services are being used? By what groups? At what levels? Are some tools and services under-utilized? Over-utilized? Why?

d. To what extent are customers satisfied with the program?

e. What are the key contextual and organizational factors that influence customers' use of the program's tools and services? What is the magnitude of those influences?

General Question 3: Are resources reasonable relative to the objectives?

a. Are the resources assigned to the various program components adequate to achieve desired objectives?

b. Is the program leveraging funds effectively? How could additional resources be leveraged?

c. Are detailed program expenditure records maintained?

General Question 4: What are initiatives that are likely to enhance program results?

a. Are there barriers that reduce awareness of, or participation in, the program? How can existing barriers be reduced or eliminated?

b. What could be done to increase the use of the program's tools and services?

c. How can the program better reach and serve non-participants? Hard-to-reach populations?

d. What are participant and non-participant recommendations for enhancements to program process and content?

e. Are there areas for improvement in the program's administrative functions (e.g., marketing, recruitment, record keeping)?

General Question 5: How can the program be modified to perform its activities at less cost and still achieve goals?

a. Which delivery channels are working well (or not working) to achieve program objectives at minimal cost? How do these delivery channels operate?

b. How can the effectiveness of the delivery channels be increased?

c. How can costs of administrative functions be reduced without adversely impacting program services?

C. Outcome Evaluation

Quantify Savings

General Question 1: How much energy and money have been saved - directly and indirectly?

a. How much energy and money were saved by participants for the entire program?

b. How much energy and money were saved by participants for individual program components/activities?

c. What are unaccounted-for "secondary" benefits (e.g., persistence, replication, delayed implementation, spin-offs)?

d. What key contextual and organizational factors are related to the achievement of energy and money savings? What is the strength of those relationships?

General Question 2: What are the non-energy benefits?

a. What were the nature and magnitude of non-energy benefits associated with the entire program?

b. What were the nature and magnitude of non-energy benefits associated with individual program components/activities?

c. What key contextual and organizational factors are related to the achievement of non-energy benefits? What is the strength of those relationships?

General Question 3: What unexpected outcomes have occurred, if any?

a. What were the nature and magnitude of any program-related results that were not intended?

b. What key contextual and organizational factors are related to the achievement of unexpected results? What is the strength of those relationships?

Market Effects or Market Transformation

General Question 1: Are targeted markets showing signs of changing?

a. Are there market changes or effects associated with the entire program (e.g., changes in business willingness or ability to produce, distribute, or service new technologies)?

b. What changes or effects are associated with individual program components/activities?

c. How has the behavior (e.g., purchase and management decision-making and practices) of targeted actors changed over the life of the program?

d. What network effects have occurred?

e. What key contextual and organizational factors are related to the achievement of market changes? What is the strength of those relationships?

General Question 2: What is progress toward desired long-term outcomes/exit strategy?

a. What are the nature and magnitude of any external replication effects that have occurred?

b. What are the nature and magnitude of any network and spin-off effects (e.g., new businesses and technologies)?

c. How effective has the program been in reducing market barriers?

General Question 3: Have sustainable markets been created?

a. Have market actors continued new practices and behaviors over time?

b. What are the effects of the program on the system specification and sales practices of market actors who received program tools or services?

c. What key contextual and organizational factors are related to the achievement of sustainable markets? What is the strength of those relationships?

D. Impact Evaluation

General Question 1: What are the verified quantified outcomes that are attributable to the program?

a. What would have caused the observed outcomes if it were not the program? What proportion of the measured outcomes were caused by the program?

b. What is the direct impact on customer awareness and knowledge that can be attributed to the program?

c. What are the energy efficiency/renewable energy actions taken by program participants compared to actions taken by non-participants?

d. What is the direct impact of the entire program on energy and money savings?

e. What is the direct impact of individual program components/activities on energy and money savings?

f. What is the direct impact of the overall program on non-energy benefits?

g. What is the direct impact of individual program components/activities on non-energy benefits?

h. What is the magnitude of replication, persistence, network, spillover, and other observed effects that can be attributed to the program?

i. What unintended results were directly caused by the program?

j. What key contextual and organizational factors are responsible for the measured net impacts? What is the strength of those causal relationships?

E. Cost-Benefit or Cost-Effectiveness Evaluation

General Question 1: What are the benefits and costs of the program's past activities?

a. What are the retrospective benefits and costs associated with the program as a whole?

b. What are the retrospective benefits and costs associated with individual program components/activities?

General Question 2: How do program benefits and costs compare to each other?

a. Are the benefits from the program greater than program and customer costs?

b. What is the benefit-to-cost ratio (using one or more different perspectives, such as "participant" or "societal")?

c. Which delivery channels are working well to achieve program objectives less expensively, and why?

Appendix 5
Procedures for Obtaining OMB Clearance to Conduct a Survey

Procedures for Obtaining OMB Clearance To Conduct a Survey

The Paperwork Reduction Act (PRA) of 1995 requires that each federal agency obtain approval from the Office of Management and Budget (OMB) before undertaking to collect information from ten or more persons, or continuing a collection for which the OMB approval and the OMB control number are about to expire. The approval process, which is popularly known as the "OMB clearance process," is extensive and time-consuming. It requires two Federal Register notices and a detailed application to OMB. The duration for the entire process can exceed six months.

At present, only three exceptions exist to this lengthy process and only one of these has relevance for general evaluations. The three exceptions are (1) surveys of Federal Government employees on subjects concerning their employment, (2) emergency reviews when a data-collection activity must be performed to meet an unanticipated, urgent need and no time is available for public comment, and (3) customer-satisfaction surveys. The first of these requires no OMB clearance. The second two require much-abbreviated processes, but of these, only the last is relevant to evaluation surveys. The last is relevant to process-evaluation surveys and the process for obtaining it is known as a "generic clearance process." Within the Department of Energy (DOE), it is sometimes called the "DOE–887 Process." The full OMB Clearance process is sometimes called the "PRA Review Process;" this appendix uses that term for the full process.

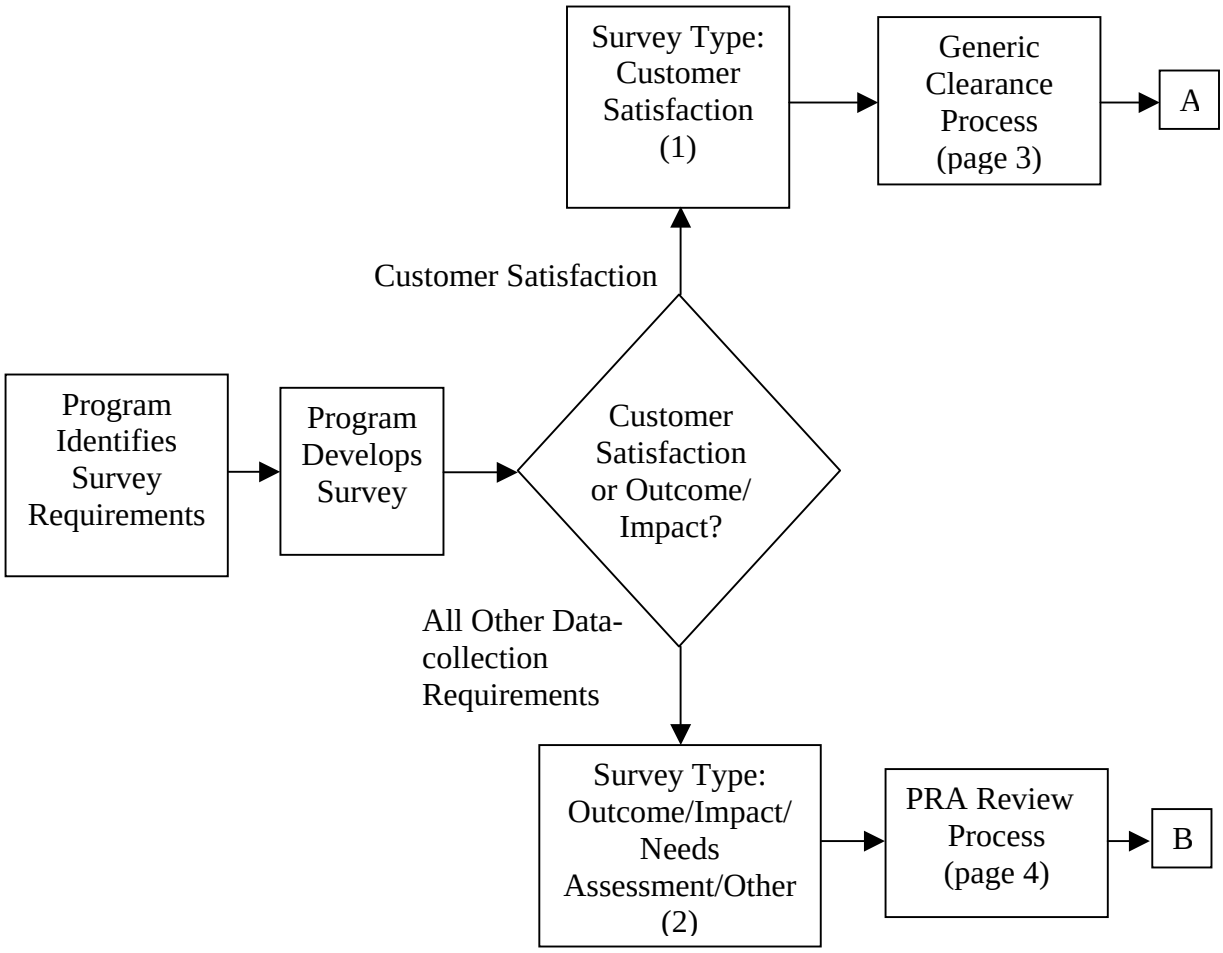
DOE's Office of the Chief Information Officer (OCIO) has issued a *Reference Guide for Submitting DOE Information Collections for OMB Review and Approval* (April 2004) (*Reference Guide*) for program managers who must collect data from the public. This 42-page document is an excellent summary of the OMB clearance requirements, including the generic clearance process requirements, and can be downloaded from: http://cio.doe.gov/RBManagement/Records/PRA_Information.html.

This appendix uses two flow charts (Generic Clearance Process and PRA Review Process) adapted from the *Reference Guide* to provide a visual image of the OMB clearance process. The flow charts show the basic decision points for selecting the appropriate clearance process and the steps for obtaining clearances under the two processes. The flow charts are based on information from the *Reference Guide* and other sources and assume data-collection from the public. The charts should be used in conjunction with the detailed guidance and contact information contained in the *Reference Guide*.

In the OMB-review flow charts that follow, each step (box) is numbered. For some of the steps, additional information about the step is provided in similarly numbered, double-bordered boxes below the step's box.

Because the PRA Review Process has many steps, an overview flowchart (page 4) showing all of the steps precedes the flow charts with additional information. This overview flowchart serves as a map to the steps of the PRA Review Process.

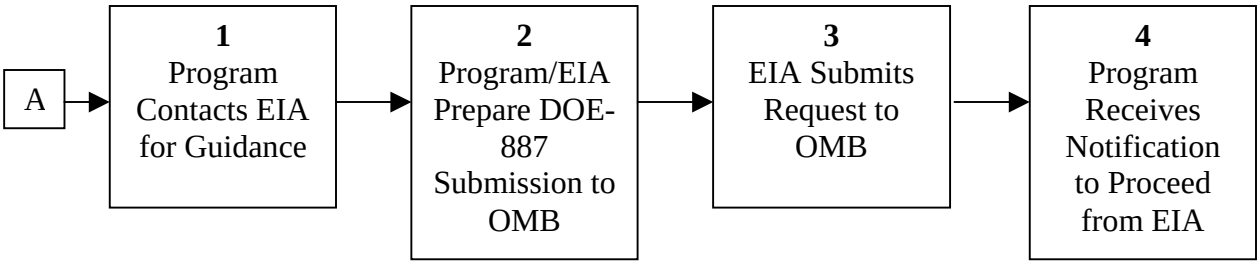
Determining Which OMB Clearance Process to Use



(1) Customer Satisfaction: Obtain information from users of EERE program outputs regarding their experience with the program and satisfaction with its products and services.

(2) Outcome/Impact/Needs Assessment/Other: Collect data for the purposes of quantifying achievements of program outputs and outcomes; assessing the proportion of outcomes that can be attributed to the program instead of other influences; estimating the needs of prospective markets that the program can meet with improvements to its design (exclusive of satisfying the needs of existing markets with the program in its current form).

The Generic Clearance Process



1

Contact EIA at 202-287-1717 for details.

4

OMB has 10-day window to disapprove or suggest changes. If OMB does not reply within this period, the clearance is approved.

2

Items to include in a letter to OMB requesting generic clearance:

- Identify organization(s) that will be using the clearance
- Reference generic clearance agreement DOE-887 "DOE Customer Surveys"
- Reference 10-day time limit for OMB disapproval
- Make commitment to provide survey results to OMB.

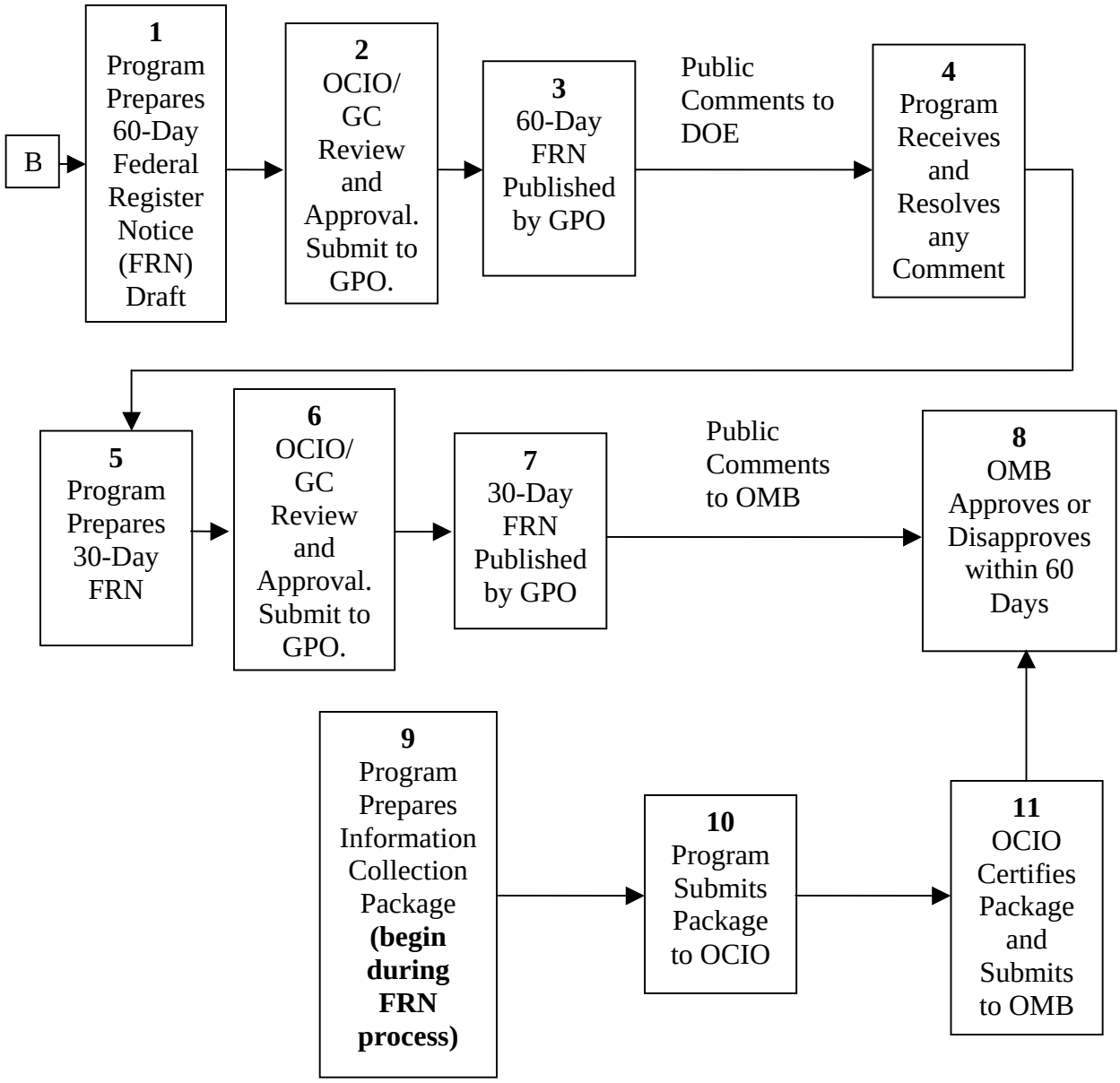
Include following information about the survey:

- Function(s) and objective(s) of requesting organization(s)
- Reasons for developing survey
- How survey responses will be used (e.g., improved program delivery)
- Description of survey respondents (what population will be surveyed?)
- Survey distribution and response collection method
- Estimate of response rate/number of respondents
- Participation factors (type of respondent, voluntary participation, data confidentiality)
- Estimate of time burden for responding for a single respondent and the total estimated burden (single response time x estimated number of responses).

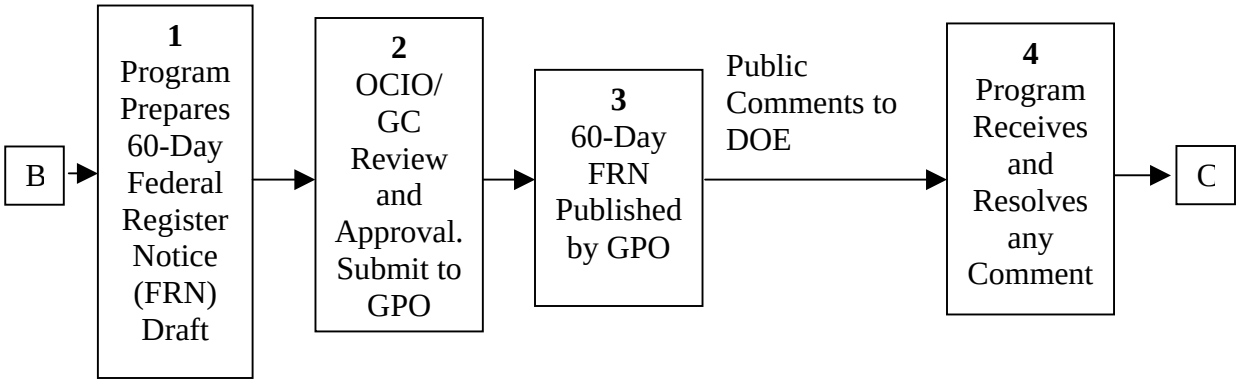
An example of a letter to OMB requesting generic clearance that has been successfully used by DOE begins on page 8 of this appendix.

The PRA Review Process – Overview

(The following three pages provide additional information on the requirements of the PRA Review Process)



The PRA Review Process – Additional Information (1)



2

Send FRN to OCIO Records Management Division. OCIO will handle GC review and submission to GPO.

4

If comments are received during the 60-day public comment period, they must be evaluated and responses must be prepared. A summary of the public comments, including those actions taken in response to the comments, must be included in the information collection package (see Box 9 on the final page of this PRA Review Process flow chart). This is particularly applicable where the proposed information collection was amended or refined based on the public comments.

1

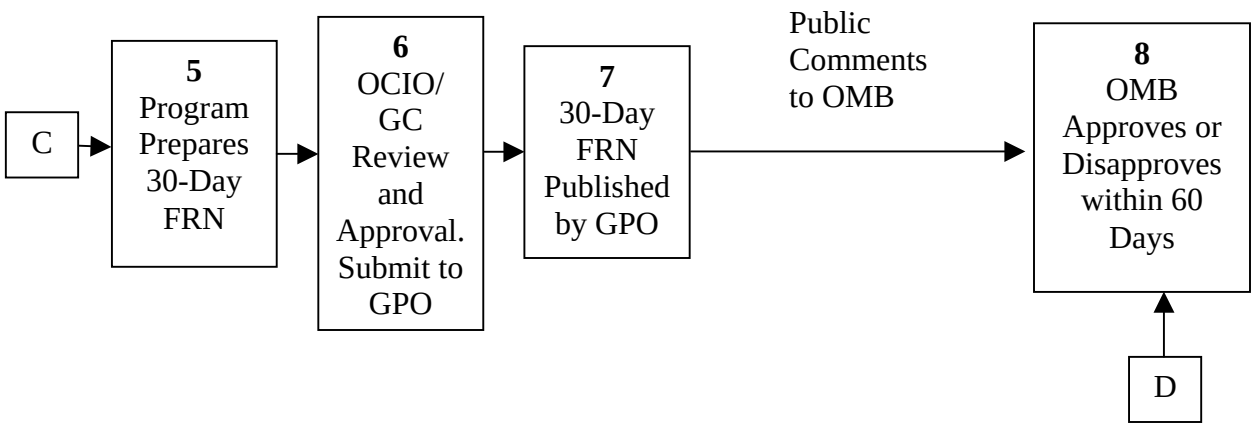
The FRN for the collection should include the following information:

- A statement that DOE is proposing an information collection to OMB
- The title of the information collection
- A summary of the information collection
- A brief description of the need for the information and its proposed use
- A description of the likely respondents and proposed frequency of response
- A summary of the privacy risks involved in collecting the information electronically from potential respondents (if appropriate)
- An estimate of the total annual reporting and record keeping burden
- Direction that comments should be submitted to DOE (see below).

The FRN should indicate that public comments and requests for information, including copies of the proposed information collection and supporting documentation, should be submitted to DOE within 60 days of the date of publication in the Federal Register. The notice should indicate that comments must be sent to the sponsoring program office along with a copy to:

Director, Records Management Division, IM-11
Office of the Chief Information Officer
Germantown Bldg., US Department of Energy,
Washington, DC, 20585-1290

The PRA Review Process – Additional Information (2)



5

After the 60-day FRN has been published in the *Federal Register* and appropriate action has been taken to address any public comments received, a 30-day FRN must be prepared to notify the public that the information collection is being submitted to OMB for review and approval.

The 30-day FRN should include the following information:

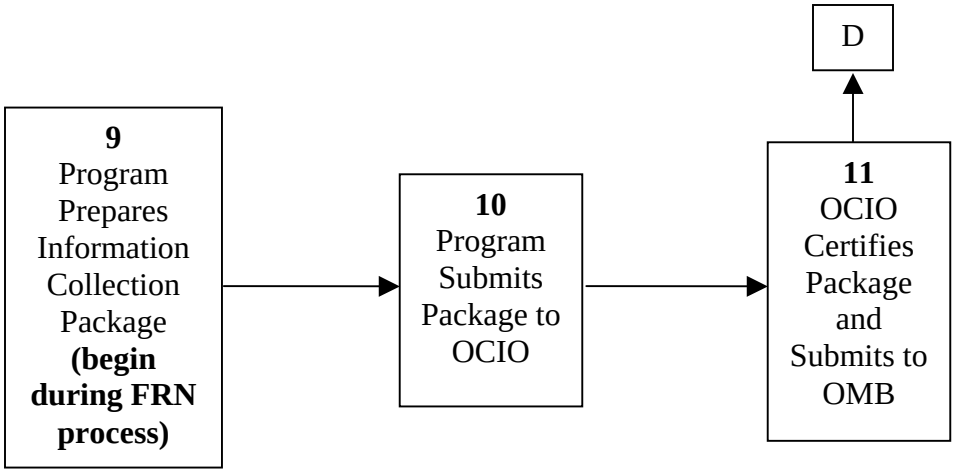
- A statement that OMB approval is being sought
- The title of the information collection
- A summary of the information collection
- A brief description of the need for the information and its proposed use
- A description of the likely respondents and proposed frequency of response
- An estimate of the total annual reporting and record keeping burden
- Indication that comments be submitted to OMB
- Statutory authority for collecting the information

The complete information package (see Box 9 of this PRA Review Process) should be submitted to the OCIO along with the 30-day FRN

6

Send FRN to OCIO Records Management Division. OCIO will handle GC review and submission to GPO.

The PRA Review Process – Additional Information (3)



9

The information collection package consists of the following documents:

- OMB Form 83-I, "Paperwork Reduction Act Submission"
- Form 83-I **Supporting Statement**
- Draft information collection tools/instruments (e.g., forms and accompanying instructions, copy of the citation from the governing regulation)
- Summary of public comments received
- Supporting documentation:
  - Copy of the signed 30-day FRN
  - Updated list (in table format) of packages contained in the collection (see Attachment VI of the *Reference Guide*)
  - Copy OR summary of the PIA (if appropriate) (See p.19 of the *Reference Guide)*
  - Other documentation deemed necessary by OMB.

The **Supporting Statement** documents that the certification requirements contained in OMB Form 83-1 have been met. See pp. 18-19 and Attachment V in the *Reference Guide* for additional information.

OMB Form 83-I requires certification that the proposed collection of information:

- Is necessary for the sponsoring program office to perform its functions
- Avoids unnecessary duplication
- Reduces the burden on the respondents
- Uses plain, coherent, and unambiguous terminology
- Will be consistent and compatible with current reporting and record-keeping practices
- Indicates the retention period for record-keeping requirements
- Informs respondents about:
  - Why the information is being collected
  - How the information will be collected
  - How the information will be used
  - The extent of the estimated labor and cost burden to respond
  - The nature of response expected (voluntary, required, or mandatory
  - The level of confidentiality to be imposed
  - The requirement that a valid OMB control number must be displayed.
- Was developed by an office that has planned and allocated resources for managing the information to be collected
- Uses statistical survey methodology
- Uses information technology to reduce burden.

EXAMPLE OF A GENERIC REQUEST LETTER USED BY DOE

Mr. xx
Department of Energy Desk Officer
Office of Information and Regulatory Affairs
Office of Management and Budget
Washington, DC 20503

SUBJECT: USE OF GENERIC CLEARANCE FOR THE REGIONAL OFFICES OF THE OFFICE OF ENERGY EFFICIENCY AND RENEWABLE ENERGY (EERE) CUSTOMER FEEDBACK SURVEY

Dear Mr. xx:

The six Regional Offices of the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE) plan to use the Office of Management and Budget (OMB) approved generic clearance, DOE-887, "DOE Customer Surveys" (OMB No. 1901-0302, expiring September 30, 2005) to survey stakeholders, partners and customers of the Regional Offices and EERE. If we do not hear that you disapprove of this request within the next ten days, as agreed to upon your approval of DOE-887, we will notify the six Regional Offices that they have authority to conduct this survey. We will provide OMB with the results of this data collection when we submit the annual report of surveys conducted under the generic clearance.

The six Regional Offices (located in Atlanta, Boston, Chicago, Denver, Philadelphia, and Seattle) of EERE are the Federal Government's principal mechanism for delivering energy efficiency and renewable energy programs at the regional, state, and local levels. For most energy users, the Regional Offices are the gateway into EERE's portfolio of technical and financial resources, designed to increase the use of efficient energy and clean power technologies in buildings, industry, transportation, power generation, and Federal facilities.

The Regional Offices have developed the attached subscriber survey in response to Executive Order 12862, "Setting Customer Service Standards." The responses to the proposed questions will be used as input for improving the delivery of EERE programs to States and local governments, communities, private businesses, non-profit organizations, colleges and universities, and other customers. The respondents of this survey will consist primarily of State government officials and other customers of the Regional Offices.

The survey will be distributed via e-mail to key partners and stakeholders in the six EERE regions from each of the six Regional Office Directors. The responses are also expected to be collected via e-mail. The key partners and stakeholders will be selected by each of the Regional Office Directors, with input from their staff. Given the interest of the partners and stakeholders in the performance of the Regional Offices and past experience, the response rate is expected to be at least 50% of the approximately 240 total partners and stakeholders receiving the survey. The intent of the survey is to obtain qualitative and quantitative information that will be useful for making improvements in the performance of the Regional Offices in delivering programs to State partners, stakeholders, and other customers. The survey results will reflect the opinions of those partners and stakeholders who choose to participate. Participation will be voluntary, and data will be confidential.

The attached survey will take approximately 15-20 minutes to answer, and contains 5 questions. The total burden for all the respondents should not exceed 40 hours (120 respondents x 20 minutes).

yy is the point-of-contact for the survey and may be contacted at zz. Other questions should be directed to aa on bb.

Sincerely,

xx
Director
Statistics and Methods Group
Energy Information Administration

Enclosure – Customer Feedback Survey

Appendix 6
Examples of General Evaluation Report Outlines[1]

[1] Some of the outlines have been edited to remove some outline levels that cover detailed project-specific results.

Examples of General Evaluation Report Outlines

Introduction

This appendix contains edited examples of the outlines of actual general evaluations performed on EERE deployment programs. The outlines were developed by the evaluators in response to the needs of the program managers and in accordance with the practice of the evaluator. Therefore, although they differ, all were judged acceptable by the respective program managers.

In spite of these differences, it is possible to identify the principal elements of a general evaluation report that should appear in every report. The following presents such a list of required elements organized as a model table of contents that contains many elements of best practice from the different examples provided. It is not necessary that these be the actual section headings; however, the report should cover the content shown.

Model Table of Contents

Table of Contents
List of Figures
List of Tables
Acronyms
Abstract
Acknowledgments

Executive Summary
- Include highlights of key recommended improvements to the program, if relevant

1. Introduction
   Program Overview (Program Description, objectives, etc.)[1]
   Evaluation Objectives and Methods (includes -- Setting for the evaluation, evaluation objective statements should be clearly specified, brief overview of evaluation methods including limitations)
   Structure of the Report

2. Study Methodology
   Data Collection Approach(es)
   Analysis Methods
   Limitations, Caveats

3. Key Evaluation Results (answers for all of the questions specified for the evaluation)
   - Could include several sections on findings
   - Findings could be presented for each different methods used, by different program components covered, by market segments covered, and so forth, followed by a section on integrated findings
     or
   - Organized and presented by the different observed effects or type of results
     or
   - Other ways

4. Recommendations (if relevant; depends on type of evaluation)
   - Clear, actionable, and prioritized recommendations that are supported by the analysis

5. Summary and Conclusions

References

(Continued)

[1] A more detailed program description could be given in a separate section, depending on the type and depth of coverage of the evaluation.

Model Table of Contents (continued)

Appendices (examples):

- Recommended improvements to the evaluation process, including any lessons learned for future evaluation studies
- Appendices containing detailed documentation of the research design and assumptions, data collection methods, evaluation analysis methodology, results tables, etc.
    - Survey or interview instrument, coding scheme, and compiled results tables and data
    - Sources and quality (caveats on data) of primary and secondary information
    - Details on quantitative data analysis: analytical framework, modeling approach, and statistical results
- Qualifications and extensions
    - Possible sources of overestimation and underestimation
    - Treatment of issues concerning double counting, use of savings factors, synergistic effects (if any), and other technical issues
    - How attribution was addressed (for impact evaluation)
    - Sensitivity of energy savings estimates
    - Assumptions and justifications

The following five outlines from actual general evaluation reports illustrate typical outlines. They do not each contain all of the elements of an evaluation report listed above. The example outlines were prepared and their reports completed before the publication of this Evaluation Management Guide.

Outline A 2001 FEMP Customer Survey

Outline B Evaluation of the Motor Challenge Program

Outline C Industrial Assessment Center Program Impact Evaluation

Outline D Preliminary Examination of Energy Management Metrics for the Best Practices Program

Outline E Building America Program Evaluation

Outline A
2001 FEMP Customer Survey[2]

Acknowledgements

Executive Summary

Table of Contents
List of Figures
List of Tables

1. Participant and Non-Participant Profiles

2. Awareness and Use of FEMP Services

3. Contact and Satisfaction with FEMP

4. Project Implementation and FEMP Influence

5. Project Needs and Possible FEMP Roles

6. FEMP ESPC Impact Issues

7. FEMP ESPC Market Issues

8. FEMP ESPC Process Issues

9. FEMP SAVEnergy Audit Impact Issues

10. FEMP SAVEnergy Market Issues

11. FEMP SAVEnergy Audit Process Issues

12. Study Methodology
    - Interviews with staff
    - Survey development process
    - Survey implementation process
    - Sampling population
        - *Sample design*
        - *Survey procedures*
    - Summary of findings

[2] TecMRKT Works and Sandia National Laboratories, "2001 FEMP Customer Survey," June 2001

Outline B
Evaluation of the Motor Challenge Program[3]

Table of Contents

[3] Xenergy, Inc., "Final Report, Evaluation of the Motor Challenge Program,", prepared for Oak Ridge National Laboratory, May 10, 2000.

Outline E
Industrial Assessment Center Program Impact Evaluation[4]

Contents

[4] Oak Ridge National Laboratory, "Industrial Assessment Center Program Impact Evaluation." ORNL-CON-473, December 1999.

Outline F
Preliminary Estimation of Energy Management Metrics for the Best Practices Program[5]

Table of Contents

Executive Summary

1. The Structure of Best Practices Energy Management

2. Structure of Energy Savings Estimates

3. Sources of Information

4. Summary of Findings
   - 4.1 Energy savings
   - 4.2 Qualifications and extensions
     - 4.2.1 Possible sources of overestimation
     - 4.2.2 Possible sources of underestimation
   - 4.3 Sensitivity of energy savings estimates; the importance of software

5. Suggestions for Subsequent Information Collection
   - 5.1 Recommendations for improved program activity tracking
   - 5.2 Recommendations for improving energy estimates
   - 5.3 Recommendations for additional data collection

6. Conclusions
   - 6.1 Summary of findings
   - 6.2 Next steps

Appendix: Assumptions and Justifications

List of Acronyms

[5] Oak Ridge National Laboratory, "Preliminary Estimation of Energy Management Metrics for the Best Practices Program," July 2002

Outline G
Building America Program Evaluation[6]

Table of Contents

Volume I. Main Report

Executive Summary

Introduction

Part I. Project Design and Overview

1. Motivation for this Project

2. Research Methods
   - 2.1 Theory and Questions
     - 2.1.1 Assessing Program Design
     - 2.1.2 Assessing Program Implementation
     - 2.1.3 Assessing Program Outcomes
   - 2.2 Methods and Data Analysis
     - 2.2.1 Qualitative Methods
     - 2.2.2 Quantitative Methods

Part II. Program Concept and Design

3. Program Concept
   - 3.1 Innovation Gap: Industry Organization and Marketplace
   - 3.2 Innovation Style: Occupational Culture and Experience-based Learning
   - 3.3 Innovation Spur: Governing Institutions and Technology Attenuation
   - 3.4 Innovation Fix: A Collaborative Public-Private Partnership

4. Program Design
   - 4.1 Team-based Organization
   - 4.2 Housing Projects and the Model for Technology Learning

Part III. Program Implementation

5. Building America Management and Teams
   - 5.1 Program Management
   - 5.2 The Teams and Their Relationships
   - 5.3 Teams Structures and Strategies

[6] Kennedy School of Government, Harvard University, Energy Technology Innovation Project (ETIP), Building America Program Evaluation, September 2004.

6. Partnership and Collaboration
   - 6.1 Collaboration in the Teams
   - 6.2 Collaboration within the Broader Partnership

7. Collaboration Effects
   - 7.1 Builder Motivation
   - 7.2 Builder Participation
   - 7.3 Builder Relationship Networks

Part IV. Program Outcomes

8. A Preface about Program Reporting, Metrics, and Data Collection
   - 8.1 Project Definition
   - 8.2 Reporting
   - 8.3 Metrics and Data Collection

9. Technology Research Outcomes
   - 9.1 Types of Projects
   - 9.2 Types of Innovations
   - 9.3 Effect of Collaboration with Builders on Technology Research Agendas

10. Technology Diffusion Outcomes
    - 10.1 Data and Findings about Builder Population and Program Experiences
    - 10.2 Findings about Technology Uptake

Part V. Lessons Learned

11. Summary and Findings about Cooperative Technology Partnership
    - 11.1 Conclusions
    - 11.2 Recommendations
    - 11.3 Lessons for Future Studies
    - 11.4 Final Thoughts

References

Volume II. Appendices

Appendix A. Building America Program Intent and Scope
- A-1. Program Overview
- A-2. Overview of Select Related Programs

Appendix B. Research Methodology
- B-1. Qualitative Research Methodology
- B-2. Quantitative Research Methodology

Appendix C. Quantitative Data Analysis: Analytical Framework, Modeling Approach, and Statistical Results
- C-1. Analytical Framework and Survey Data Summary
- C-2. Categorization of Variables
- C-3. Linear Regression Model Development
- C-4. Regression Output Tables

Appendix D. Builder Survey Instrument, Coding Scheme, and Compiled Data

Appendix E. Overview of Teams, Contracts, and Funding

Appendix F. Team Reporting and Project Summary Examples

Appendix G. Program Recommendations

Appendix 7
EERE Quality Assurance Guidelines for General Program Evaluation Studies

EERE Quality Assurance Guidelines for General Program Evaluation Studies

Quality assurance (QA) is essential to the effective development of general program evaluations if they are to have high credibility with program managers, management, OMB, Congress, and other stakeholders. Quality is more than meeting a set of standards, it is an ingrained attitude that plays out in day-to-day behaviors that result in the delivery of programs and products that meet and exceed the needs and expectations of customers and stakeholders.

EERE expects that program managers will hire third-party expert reviewers to review at least the Evaluation Plan and draft report(s). How programs implement such third party review will depend on the scale and the scope of the evaluation. Table 1 provides guidance.

However the review is implemented, it is important to provide the reviewers with a scope for the review. The reviewer should be asked to address a specific set of factors designed to guide but not limit the reviewer's input. The scope for the review is especially important when multiple reviewers are involved. Reviewers should feel free to raise issues that are outside the scope for the review. This appendix provides a list of such factors.

Table 1: Options for Implementing a Third Party Quality Assurance Review Panel

Evaluation	QA Review Panel
Evaluations with larger budgets, that are broad in scope, that have wide stakeholder interest, and that may be difficult to implement	Establish a standing peer review panel comprised of 4-5 outside experts to review the Evaluation Plan and the draft report. The standing peer review group would attend two or three one-day on-site meetings and provide written comments. The evaluation project manager should respond to review comments and modify the Evaluation Plan and draft report as appropriate.
Evaluations with small budgets and narrow scope	Choose two or three outside reviewers with appropriate skills, provide them with the documents and information to be reviewed, and ask them to provide written feedback. The evaluation project manager can: 1) utilize the feedback directly with no further input 2) circulate the feed back among the reviewers and convene a teleconference to discuss the results. The latter option may be the preferred option when there are differences among the reviewers.
Selected technical aspects of an evaluation, such as sampling plans, questionnaires, guides, or data analysis	Select two or three persons who understand the content and the technical requirements and have them provide written suggestions.

The table that follows lists the key evaluation activities that should benefit from a QA review and identifies the related QA factors. The table also identifies the categories of individuals that should be involved in review at each stage. The factors listed in this table can be used to construct a QA plan for specific general program evaluations.

Table 2: Quality Assurance Factors and QA Review Participants

Quality Assurance Factors	Participants in the QA Review
Development of the SOW	
• Includes statement of objectives of the evaluation. • Focus is on issues that are critical to the program. • Evaluation has the backing of program managers, managers, stakeholders and others who have an interest in the outcome of the evaluation. • Utilizes evaluation designs and data collection and analysis techniques that are appropriate to the goals and objectives being addressed. • Utilizes evaluation designs and data collection and analysis techniques that are based on generally accepted scientific evaluation practice. • Has funding and timelines that are commensurate with the expected results. • Incorporates procedures that test to see that data collection activities are reliably producing valid data. • Provides for continuing review of intermediate, draft and final work products. • Program to be evaluated is described sufficiently well to allow evaluators to understand it. • Key questions to be answered by the evaluation have been identified. • Proposed data collection and analysis methods will produce relevant data at an acceptable cost. • Data collection and analysis methods are feasible. • Resources and timeframe in which the evaluation is to be conducted are sufficient to complete the research activities.	• Evaluation project manager • Program manager • Stakeholders interested in the results of the evaluation

(Continued)

Quality Assurance Factors	Participants in the QA Review
Contractor Evaluation Plan	
• QA panel consisting of one or more third party evaluators has been established. • Contractor understands the program. • All of the key evaluation questions have been identified. • Key questions can be answered with the proposed data collection and analysis methods. • Proposed sample sizes are adequate. • Procedures for drawing the sample frames and collecting the data are feasible. • Those procedures will produce data that are relevant to the questions to be answered. • Procedures to assure the validity and the reliability of the data to be collected have been proposed. • Analysis methods are appropriate and contractor has the knowledge and skill to undertake the analysis. • For statistical methods, the degree of relationship between indicators, tests of significance, and confidence intervals for sample estimates will be included in the analysis, as appropriate • The analysis will incorporate results of any previous related studies. • There is a reasonably detailed outline or description of the expected reports or products that matches the goals of the evaluation. • Contractor work plan has been submitted for review to the QA panel.	• Evaluation project manager • Program manager • Others as identified by the evaluation project manager • QA panel (see Table 1)

(Continued)

Quality Assurance Factors	Participants in the QA Review
Sampling, Data Collection Instruments, and Data Collection Activities (Either as described in the Contractor's Evaluation Plan or as Produced during the Project)	
• Contents of the data collection instruments map to the researchable issues and goals and objectives of the evaluation. • There is sufficient sociographic, firmographic or demographic data to describe the populations of interest. • Any guides or survey instruments are structured in accordance with good design principles. • In surveys, closed response sets have been used wherever possible and the items in the response sets are complete and unique. • Guides or questionnaires been reviewed by people who understand the program. • Surveys have been pre-tested. • Data validation procedures have been incorporated. • Population list or sample frame represents the population. • There are procedures in place to replace sample points for refusals, dropouts, and points that cannot be contacted. • Any inferred data values use appropriate inference methods. • Above factors are examined by QA panel during the Evaluation Plan review.	• Evaluation project manager • Program manager and/or program personnel • QA panel (see Table 1)

(Continued)

Quality Assurance Factors	Participants in the QA Review
Draft Report	
• QA panel consisting of one or more third party evaluators has been established. • There is an approved outline for the report. • The outline for the report includes a description of the program, a list of the researchable issues, a discussion of data collection and analysis methods, a discussion of the limitations of any data collected, and a logically organized set of content chapters. • `The data have been structured properly for the analysis. • The metrics and indices that have been constructed have a substantive, logical and statistical basis. • The analysis techniques are appropriate. • The discussion in the text is supported by the data. • The summaries and conclusions follow from the analysis. • The reports present answers to all questions asked. • Report versions have been submitted for review to the QA panel or equivalent.	• Evaluation project manager • Program manager • Program personnel • Stakeholders • QA panel (see Table 1)
On-going	
• If the evaluators have encountered difficulties that will impact schedule or budget there is a plan for addressing these difficulties	• Evaluation program manager • Evaluation contractor

Appendix 8
Outline of a Model Statement of Work for an Evaluation Study

Outline of a Model Statement of Work for an Evaluation Study[1]

The statement of work (SOW) for a general program evaluation consists of a description of the objectives of the evaluation and specifications on how the objectives should be achieved. The description and specifications should have enough details to allow a prospective evaluation contractor to prepare a convincing proposal demonstrating that it can achieve the objectives. The model outline of a SOW in this appendix is intended as a guide for EERE staff who must prepare an SOW for a general evaluation study. It is a generic outline; it is not customized to a specific type of evaluation, i.e., market, process, outcome, impact, or cost-benefit evaluation.

A SOW used by FEMP for a preliminary study of the metrics to use for outcome and impact evaluations of FEMP's Technical Assistance Program is attached to this appendix as an example of a completed SOW for an evaluation study.

A SOW performs the following functions:

- It describes EERE's expectations for the evaluation.
- It describes the key elements of a planned activity or analytical effort that EERE expects the evaluation contractor to perform.
- It is used by EERE program staff to develop an RFP.
- After the proposals are submitted, it should be used to inform evaluator selection.
- After an evaluator is selected, it becomes the basis for any subsequent negotiations needed to create the terms of a mutually acceptable Evaluation Plan (see Appendix A10, "Model Evaluation Plan"). (The evaluation contractor will prepare the final Evaluation Plan resulting from the SOW.)

The SOW can be prepared by:

- A professional evaluator for acceptance and approval by the EERE staff who commissions the study
- An EERE staff person
- An EERE staff person and an evaluator collaborating to establish the framework for an evaluation. This might be the case if the evaluation were to be performed under an existing task-order contract.

A good SOW, developed at the start of the evaluation project and clearly setting out the objectives, rationale, and expectations of the evaluation study, will greatly enhance the quality and usefulness of the final evaluation product. The following are the essential elements of a good SOW:

[1] This SOW guidance is based on several widely used technical notes and examples of Statements of Work or Terms of References (TOR) developed by organizations such as: UNICEF Evaluation Office, "Evaluation Technical Notes No. 2," April 2002 (www.unicef.org/evaluatoin/TechNote2_TOR.pdf); World Bank Operations Evaluation Department, (www.worldbank.org/oed/); and "Model TOR," in "How to Perform Evaluations," Canadian International Development Agency, March 2001 (www.acdi-cida.gc.ca/cida_ind.nsf/).

Evaluation Title

Brief Program Background and Context

- Describe the program history and goals.
- Describe the program's current status.
- Describe anticipated changes in the program and the reasons for them.
- Identify interactions between the program and other EERE or Federal/State programs or policies.
- Identify the key stakeholders and partners involved in the program.

Purpose and Objective(s) of the Evaluation

- Describe the need for the evaluation. What are the intended uses for decision-making?
- For whom will the evaluation be performed (what audiences)?
- Describe the objectives of the evaluation.
- Describe what will not be addressed, if this will better define the objectives by ruling out scope that might appear to be related.

Key Evaluation Questions to Satisfy the Objectives

- Identify the major (general) evaluation questions.
- Identify specific evaluation questions or provide examples (subject to further clarification of them by EERE manager and Evaluation Contractor.)

Evaluation Approach and Method(s)

- Provide an overview of how the evaluation is to be conducted. Describe the expected data collection and analysis methodologies for the evaluation.
- The approach and method may become the subject of negotiation before an Evaluation Plan is written.
- Discuss anticipated data and methodological issues and how they could be addressed. (See Appendix 9, "Lessons Learned for Improving the Quality of EERE Evaluation Studies.")

Evaluation Work Structure and Provisional Timetable

- Define the area and population to be considered, national/regional, etc.
- Define the period of program performance to be evaluated.
- Specify the start date for the evaluation and the date by which a final report is required.
- Identify specific tasks (e.g., task 1, task 2, task 3), or define tasks that organize the work into an efficient structure by which the evaluation can be managed and monitored.

- Specify expected meetings (EERE briefings, stakeholder interaction, etc.),
- Identify the scheduled reviews by outside experts (see QA Procedures).

Assistance to be provided by EERE to the Evaluation Contractor

- Include a list of documents to be provided, e.g., program records.
- List known relevant reports to be provided.
- Identify how to contact program staff, if appropriate.
- Provide assistance in obtaining an OMB clearance for a survey, if appropriate.

Products Expected from the Evaluation Contractor

- List the expected products to be delivered, to whom and when (e.g., draft and final Evaluation Plan, draft and final reports, data sets, etc.)
- Specify the initial expectations for content of the study reports, e.g., initial report outline and list of what the report should include.
- Specify the number of copies of delivered products, and who will do the publication.

Quality Assurance (QA) Procedure

- Describe QA procedures defined for the study (preferably based on EERE QA expectations and any additional procedures proposed by the contractor). See Appendix A7, "EERE Quality Assurance Guidance for General Evaluation Studies."
- Identify the roles and responsibilities of the evaluation QA team (or committee). The QA team should consist of evaluation experts who are not part of the study team.
- Specify the required composition of the QA team (subject knowledge coverage, expected qualifications). (Responsibility for identifying the QA team will be EERE's.)
- Identify the milestones during the evaluation process for participation and review of products by the QA team (e.g., beginning, review draft Evaluation Plan, and review draft report).

Organization and Management

- Identify the EERE staff contacts for questions about the SOW and contract.
- Specify other implementation arrangements between evaluators and EERE (i.e., role of EERE in evaluation data collections, etc.).
- Specify the number of trips and location of anticipated travel, if appropriate.

Resources *[For EERE internal Use only; remove this part from the SOW before any external communication of it.]*

- Projected cost and breakdown by task/activity, professional fees, travel, etc.

ATTACHMENT TO APPENDIX 8: SAMPLE STATEMENT OF WORK

DEVELOPMENT OF PRELIMINARY METRICS FOR THE TECHNICAL ASSISTANCE PROGRAM OF THE FEDERAL ENERGY MANAGEMENT PROGRAM

BACKGROUND

The mission of the Federal Energy Management Program (FEMP) is to help the United States government reduce its energy consumption. The federal government is the world's largest energy user. Substantial amounts of money could be saved through energy-efficiency investments. Additionally, because the federal government is the world's largest purchaser of goods and services, its procurement policies have the potential to transform markets for energy-efficient products, renewable technologies, and other new, energy-related technologies.

OBJECTIVES

The objectives of this project are to (1) develop a metrics framework focused on energy savings for the Technical Assistance (TA) component of FEMP and (2) develop preliminary energy savings metrics for TA. Results of this evaluation project will be used for project management purposes and will be used as inputs into FEMP's Government Performance and Results Act (GPRA) process.

KEY EVALUATION QUESTION

The key evaluation question is this: how much federal government energy savings in FY03 can be attributed to FEMP's TA program?

EVALUATION APPROACH AND METHODS

This evaluation project has been broken into the following tasks:

Task 1: Develop a framework to organize the evaluation. A matrix will represent the framework. Each row of the matrix will represent a TA program channel that could lead to energy savings (*e.g.*, direct technical assistance, energy assessments, software). Each major column of the matrix will represent a TA sub-program area (*e.g.*, ALERT, SAVEnergy, Technical Assistance/Design Assistance (TA/DA)). Each major column will be further subdivided into building types: standard buildings and energy intensive buildings. Each cell of the matrix will contain an estimation of energy savings attributable to a TA subprogram for specific building type attributable to a specific TA program channel.

A strength of this framework approach is that it links identifiable and countable FEMP activities directly to energy savings. Another strength is that the matrix represents through its columns exactly how FEMP is organized. Thus, by filling in

the cells of the matrix, FEMP will be able to assess the effectiveness of specific delivery channels and energy savings attributable to its specific programs. It should be noted, however, that there are other models that can be used to conceptualize the influence FEMP has on energy use in the federal government. For example, it is possible to conceptualize FEMP's activities as influencing the various stages of the life of a building, from its design to its construction to its operation and maintenance.

The project team will work with DOE and FEMP to relate the framework to the underlying logic of the program (as per FEMP TA logic models). In addition, as requested, the project team will work with DOE and FEMP to map the particulars of the matrix to these more general program area categories if DOE and FEMP desire estimates of energy savings associated with building life cycles.

Task 2: Identify existing data sources to estimate outputs associated with each cell of the matrix. Outputs are related to channel activities, such as the number of TA/DA projects completed, the number of assessments conducted, and the number of software packages downloaded from the website. FEMP Central is expected to be a major source of data for this project.

Task 3: Identify existing data sources to estimate energy savings outcomes associated with the estimated outputs. For some cells in the matrix, data sources probably already exist that associate outcomes with outputs (*e.g.*, FEMP Central has energy savings outcome estimates for many TA/DA and SAVEnergy projects). Those cells in the matrix for which energy savings outcome data are not readily available will be identified.

Task 4: Prioritize cells without readily available outcomes data. It may not be possible given the time and funding limitations of this project to estimate outcomes for each cell of the matrix. Therefore, efforts need to focus on those cells that are expected to yield the highest energy savings estimates. The project team will work with FEMP staff on this prioritization task.

Task 5: Develop primary data collection plan. Based on the outcomes of Task 4, a primary data collection plan will be developed to estimate energy savings for cells with high priorities. It is expected that the plan will include a survey. The plan will specify the potential survey respondents (*e.g.*, FEMP staff, project managers, and/or FEMP customers), how the respondents will be chosen, and what type of survey will be administered. The type and scope of the survey will be driven by primary data collection needs and constrained by the time and funding available to this project.

Task 6: Develop methods to estimate energy savings outcomes for those cells where outcome data are not available and where primary data collection will not be conducted. Based on other outcome evaluations conducted by the evaluation contractor, existing literature and information collected about other deployment programs probably contains many useful results that could be generalized to the FEMP context. Description of methods will clearly outline approaches to be used to assess attribution of outcomes.

Task 7: An external review of the framework and data collection/analysis plan (i.e., Evaluation Plan) will be conducted. Time will be allocated in the project schedule to allow the overall evaluation plan to be peer reviewed. A preliminary evaluation plan that provides a sufficiently detailed description of Tasks 1 through 6 will be prepared and

presented (could be in the form of a PowerPoint) to FEMP staff and small group of external reviewers.

Task 8: Collect available output and outcome data. This project will rely on existing data sources, such as FEMP Central. Organizations that may have relevant output and outcome data will be contacted and requests for data will be submitted. The primary data plan developed under Task 5 will be implemented.

Task 9: Implement the approaches developed under Task 6 to estimate energy savings outcomes for the other cells in the matrix.

Task 10: Develop a discussion about how much of the energy savings outcomes in the matrix can be directly attributable to FEMP's TA program.

EVALUATION WORK STRUCTURE

This project will address energy savings attributable to FEMP's TA program in FY03. The project will begin in July 2004. A first draft of the framework will be delivered to FEMP in July 2004. A document more clearly defining what can be accomplished by this first, preliminary metrics project within the time and funding constraints of this project, and a preliminary primary data collection plan will be delivered to FEMP in November. That document will also form the basis for material used in the external peer review to be conducted in early December. Primary data collection will begin in January 2005. A draft report will be delivered to FEMP in April 2005 to be reviewed by FEMP staff and the external reviewers. A final report will be delivered to FEMP in June 2005.

ASSISTANCE TO BE PROVIDED BY FEMP TO THE EVALUATION CONTRACTOR

FEMP will assist the evaluation contractor in its efforts to identify data resources and collect relevant data. For example, FEMP will provide the evaluation contractor access to FEMP Central. FEMP will direct the evaluation contractor to relevant data sources and will provide contact information.

DELIVERABLES

The evaluation contractor will provide the deliverables at the times listed under Evaluation Work Structure.

QUALITY ASSURANCE

All deliverables will be internally reviewed by the evaluation contractor and will be reviewed by FEMP staff. A small number of external reviewers will review and provide comments on the preliminary Evaluation Plan (see Task 7). They will also review the draft report.

ORGANIZATION AND MANAGEMENT

The evaluation contractor will provide the deliverables at the times listed under Evaluation Work Structure. The evaluation contractor staff will regularly report project progress to FEMP staff. The evaluation contractor will present the preliminary evaluation plan for the project to FEMP staff in Washington, DC in December 2004 (or at a more convenient date as necessary).

Appendix 9
Lessons Learned for Improving the Quality of EERE Evaluation Studies

Lessons Learned for Improving the Quality of EERE Evaluation Studies

Introduction

A number of lessons have been learned from critiques of past EERE evaluation studies.[1] Awareness of these lessons can help promote continuous improvement in the planning, design, and conduct of evaluation studies in EERE. It is recommended that program managers incorporate these lessons, as appropriate, into the statement of work used to hire an evaluation contractor.

Formulation of the Evaluation Statement of Work

1. ***Develop a Statement of Work (SOW) for the Evaluation Study:*** Typically, EERE general program evaluations are initiated without preparing a full SOW. This often leads to an unproductive evaluation and wasted managerial time because a full consideration of the scope of the evaluation is not established *before hiring a contractor*.

> Program staff should develop a SOW to use to hire an evaluation contractor. See Appendix 8 for a model statement of work and a specific example.

2. ***Evaluation objective statements should be clearly specified:*** Evaluation SOWs do not always describe the intended uses of the evaluation, the decisions under consideration, the types of information required, or even clearly define the evaluation objectives. The evaluation should be designed with specific objectives in mind, and these should be clearly described in the SOW.

> Program staff initially, and then in consultation with the evaluation contractor, need to clarify intended uses of the evaluation, decisions under consideration, kinds of information required, and use this information to define clear evaluation objectives.

Credibility of Results

3. ***Double counting:***

- The overlapping and interactive structure of program components can lead to possible double counting of energy savings when savings estimates attributable to each program component (or activity) are developed separately.
- EERE deployment programs may use the outputs of EERE R&D programs. In such a case both programs may claim credit for energy savings resulting from their efforts.

[1] Much of this material is assembled from EERE evaluation reports or from summaries of comments made by external reviewers at evaluation study peer review meetings.

For outcome, impact, and cost-benefit evaluations, evaluation contractors should be asked to identify areas where double counting is possible and describe how double counting would be avoided, addressed, and documented in the report.

4. ***Sources of overestimation & underestimation:*** Often, outcome or impact evaluation studies report that their estimates are "conservative" in that overestimation is outweighed by underestimation. In other cases, spillover benefits from program outcomes may be hypothesized but not quantified because of the difficulty of making reliable estimates.

For outcome and impact evaluations, evaluation contractors should be asked to clearly identify in the Evaluation Plan, and document in the report, all sources of overestimation & underestimation. Hypothesized spillover benefits should be discussed even if they are not quantified.

5. ***Use of "savings factors" in lieu of site-specific measurement:*** When savings factors, e.g., kWh saved per energy efficiency measure outcome, are used in lieu of direct measurement they must be applied appropriately to match the profile of the population that they are intended to represent. It generally will not be correct to transfer savings factors to entities that have widely different profiles compared to those from which the savings factors were derived.

Evaluation contractors should be asked to fully describe the planned methodology for use of savings factors in the Evaluation Plan, including how they intend to account for site-by-site variation, applications variation, and other variations in the profile of the study population where these factors could be significant. Where savings factors are used, develop a means to check the reasonableness of the resultant energy savings numbers across the study population (e.g., acquire and evaluate information that can be used as a benchmark).

6. ***Construction of attribution questions in surveys*:** When survey-based questions are used to address attribution, the questions have to be carefully structured to get at the attribution issue at hand. Failure to properly structure the questions will result in unreliable recipient responses. For example, a question such as "Did it influence your decision—Yes or No?" is inadequate for addressing attribution. An attribution question should not force a "yes" or "no" response. Instead, it should distinguish response by degree of influence (e.g., very little, somewhat, significant, dominant; or a numeric degree-of-influence scale).

Survey-based attribution questions in draft survey instruments should allow for the many factors that can influence choice and be reviewed by evaluation peers before the survey is fielded.

7. ***Survey non-response:*** A common problem encountered in survey work is non-response. Non-response can introduce error into survey results. The degree to which the results represent the intended population critically depends on the response rate. A poor response rate can undermine the external validity of the survey results.

> Evaluation contractors who plan to use survey research should be asked to describe in the SOW and the Evaluation Plan their approach for avoiding, minimizing, or controlling potential non-response error. In the final report they should describe how they addressed non-response, and any implications for the reliability of the results. Evaluators should not consider the non-response problem for the first time *after* the survey is fielded.

8. ***Explicit documentation of the source(s) of energy savings:*** Frequently, studies that are not based on site measurement of savings fail to clearly describe the source of their reported energy savings. Savings based on factors used by different sources, e.g., states, are provided without describing the assumptions underlying the savings factors.

> Evaluation contractors should explicitly address in the Evaluation Plan how they intend to estimate energy savings and the assumptions underlying their estimates. This should also be documented in the final report.

9. ***Describing caveats on data used in the evaluation:*** Budget constraints sometimes force compromises in the methodology used for data collection, yet the potential weaknesses created by these necessary choices are not acknowledged. The study needs to be sure to fully describe the caveats and other issues concerning the data used in the study.

> The report outline developed by EERE staff and the evaluation contractor should include a section on limitations and caveats regarding the data. The report should adequately and appropriately highlight any concerns and limitations about the data used. *Data caveats should also be mentioned in the Executive Summary for the less reliable findings and recommendations.*

10. ***Sources of information:*** Previous evaluation reports have not always described sources of data in sufficient detail to allow an independent determination of the appropriateness of the information.

> The evaluation study scope of work should stipulate that the evaluation contractor must describe sources of data in enough detail to allow the appropriateness of the data to be determined. This description should be included in both the Evaluation Plan and the Final Report.

11. ***Estimating leverage impact:*** For programs that provide funding to support investments in project areas, it is common for evaluation studies to attempt to estimate how much additional funding in project areas was leveraged by a program dollar (e.g., "one dollar in program dollars leveraged xx million dollars in additional funding"). These leverage impact estimates are sometimes grossly exaggerated. A common problem of evaluation studies when determining the leverage effect is that they do not always adequately address attribution or account for the nature of financing in the subject project areas.

> For studies that attempt to estimate and report leverage impact it is essential to determine the extent to which the "non-program players" also devote dollars to program-targeted project areas independently of program funding (e.g., funds from System Benefit Funding sources). Also, one must determine the amount of funds in the project areas that program beneficiaries would have invested even if the program funds were not available. Absent this and other information about the project financing, it will be difficult to know who is leveraging whom.

Interactions within Program and across Programs

12. ***Synergistic effects among program elements:*** Studies do not always make an effort to assess the synergistic effects among program elements – e.g., how a combination of publications, software tools, and technical assistance might be more effective than each as a separate entity.

> As appropriate, evaluation contractors should be asked to describe in the Evaluation Plan how they intend to assess the synergistic effects among program elements. However, avoid double counting. (See item #3.)

13. ***The same population receives the services of multiple programs***. For example, how do deployment activities and other programs that provide direct service to the same set of customers interact to produce a customer choice? How should the resulting outcomes be allocated?

> Program staff should clearly document what other programs within or outside of EERE also serve their program's target audience. For impact evaluations, the Evaluation Plan should include a discussion of this issue and the plan for addressing it.

14. Accounting for "shelf life" of programs' products: Energy efficiency measures and practices do not last forever. The effectiveness of most energy-efficient measures deteriorates with time. All have a useful effective life. These effects should be applied to the benefits side of cost-benefit evaluations.

> EERE staff and the evaluation contractor should decide how to account for savings shelf life. The evaluation contractor should describe in the Evaluation Plan how this will be accomplished.

Findings and Recommendations Presented in Reports

15. Precision of reporting of the results: Reports sometimes report results at a level of precision that is not justified by the data and analysis.

> Evaluation contractors should not report numbers with too many decimal places. In some cases, the evaluation contractor might consider reporting results as a point estimate within a range.

16. Provide a list of clear, actionable and prioritized recommendations that are supported by the analysis: Some evaluation studies have not developed program-improvement recommendations for the client to consider, or do not always develop recommendations that are adequately supported by the analysis. Similarly, recommendations for improving the quality of the evaluation are often omitted, even though the evaluation report acknowledges difficulties in performing the evaluation.

> Evaluation contractors should be asked to provide an explicit set of recommendations for both program and evaluation improvement, as appropriate, and ensure they are supported by the analysis conducted. Recommendations should be ranked in priority order (high, medium, low).

17. Rank findings by level of defensibility: Outcome and impact evaluations that estimate savings by component or activity levels typically do not associate a level of defensibility to each reported component result.

> For outcome or impact evaluations, evaluation contractors should report on the level of defensibility of each estimate associated with a particular program component for which a quantified finding was developed. This need not be a quantitative value; a subjective ranking should be feasible based on the relative strengths and weaknesses of the respective methodologies. An alternative approach to this would describe caveats for the findings as described under **Credibility of Results** above.

18. Program record keeping and database recommendations: Program record keeping and databases are rarely designed to support evaluation activity. Often information about participants that is important for evaluation procedures is missing from program records.

> Evaluation contractors should make their program record-keeping recommendations for general program evaluation purposes explicit so the program can begin to collect these data for future evaluations.

Appendix 10
Model Evaluation Plan Outline

Model Evaluation Plan Outline

Introduction

The Evaluation Plan establishes the written plan for conducting a general program evaluation. The evaluator should prepare an Evaluation Plan and obtain the program manager's written approval of it before investing any of the study's resources in the actual evaluation. The program manager should have it reviewed by the evaluation quality assurance team (see Appendix 7) before approving it.

The plan provides the manager and the evaluator a mutually agreed-upon understanding of how the evaluation will be performed and what its output will be. As such, it provides the program manager with an important tool for monitoring the evaluator's progress and managing the project.

In practice, the Evaluation Plan is often developed directly from the evaluator's proposal; therefore, the outline in this appendix has many similarities with the model outline for an evaluation statement of work in Appendix 8. The principal difference in the two is their perspective: the statement of work reflects what the manager wants the evaluator to propose to do. The Evaluation Plan reflects the actual performance plan after the manager and the evaluator have reached agreement on:

- Any changes in the manager's requirements since the statement of work was prepared
- The answers to any questions the manager or the evaluator has at the outset of the project
- A performance schedule based on the actual start date for the study
- Dates for delivery of any assistance required from EERE
- The resolution of any other issues that either party has regarding the required and proposed statements of work.

The model Evaluation Plan outline in this appendix is typical but not inflexible. Its elements may be grouped differently or placed in a different order. If an evaluation study will perform only part of a complete evaluation (see, for example, the attachment to Appendix 8), elements can be omitted. However, each of the elements in this model outline that are noted with an asterisk (*) should be included in an Evaluation Plan, regardless of its scope.

Outline for a Typical Evaluation Plan[1]

9. Table of Contents

List of Tables and Figures

1. *Introduction

2. *Program Background

This section should provide the following:

- *Provide an overview of the program, including its objectives, and the activities and outputs designed to produce its outcomes. Include detail on those activities and outputs that the study should evaluate.
- *Describe the decisions about the program for which the evaluation will provide input.
- *List any program materials that can contribute to the evaluation of the program. Describe their availability.

3. *Objectives

- *Describe the objectives of the evaluation. Tie them to the decisions to be made about the program.
- *Describe the type of evaluation(s) needed to satisfy the objective(s).

4. Logic Model

- Provide the current logic model of the program. (If the evaluator must develop this, as part of the evaluation, this section can be omitted, but the Work Plan should include a task to develop a logic model.)

5. *Evaluation Questions

- *List the general questions that need to be asked in order to meet the objectives of the study.
- *List the specific questions to be answered in order to answer the general questions.
- If a logic model has been provided, the general and specific questions must relate to inputs, activities, outputs, outcomes, or external influences that are depicted in the logic model.
- If the evaluator will develop the logic model, the general and specific questions must be related to its inputs, activities, outputs, outcomes, or external influences.

[1] Items with an asterisk (*) should be included in the Evaluation Plan.

6. Overview of Evaluation Approach

Provide an "executive summary" of the research design (if appropriate), data collection approach, and analytical approach. This might include a workflow diagram.

7. *Work Plan

This section describes the steps by which the evaluator will satisfy the evaluation objectives. It describes the work elements, the research design (if appropriate), the data collection plan, the analysis plan, the deliverables, the evaluator's expectations for assistance from EERE, and a schedule for performance.

The evaluator should reference steps in the Work Plan in making its work progress reports.

7.1 *Work Structure

This section describes the task structure for the work to be performed. This is the part of the Evaluation Plan against which the program manager should track and manage the evaluator's performance and compensation.

The following is a typical task structure. Individual studies may vary from this, according to their work requirements and the program manager's preference for tracking contract work. Tables and diagrams are encouraged to clarify or simplify elements of the work structure.

- Task 1: Project Management
  - *Specify a Project Initiation Meeting, including requirements for minutes of the meeting.
  - *Develop the Project Evaluation Plan. (This element is *in* the Evaluation Plan so that it may be reported as completion of a project requirement.)
  - *Specify the requirements for progress reporting.
  - *Specify the frequency of meetings between the project manager and the evaluator for project progress and management.
- Task 2: Develop the Research Plan
  - *Describe the overall approach to the research if this has not been described in a prior section, such as Section 5 of this outline.
  - If an impact evaluation is being performed, describe the research design for developing valid and generalizable findings. Discuss the internal and external validity and the relative defensibility of the design.
- Task 3: Data Collection Plan

(The actual data collection steps may be separate tasks or part of an overall data collection plan.)

- *Specify what data will be collected to answer the general and specific questions.
- *Specify the type of data collection activity, e.g., a survey.
- *Specify the population and how the subjects will be selected for data collection.
- If data will be collected by a sample, specify the source of population data that will constitute the sample frame.
- If data will be collected by a sample, specify how many respondents will be contacted and the design statistical precision, if random selection procedures will be used.
- If OMB clearance is required, the effort for this may be included in the Data Collection Plan or be a separate task.

- Task 4: Data Analysis Plan

  (This element describes how the evaluation findings will be produced from the collected data. If several different analytical methods will be used, they may be separate tasks or part of an overall analysis plan.)

  - *Specify the analysis approach.
  - *Specify what findings the analysis(es) will produce.

- Task 5: Reporting
  - *List the draft and final deliverables needed to satisfy the evaluation objectives.
  - Provide a preliminary outline(s) of the principal report(s).
  - Specify the number of copies of each deliverable to be provided, and the stakeholders who will receive the copies.
  - *Specify the content, and frequency of the progress reports.
  - Specify the raw or cleaned data that must be delivered to EERE.

7.2 Assistance to Be Provided by EERE

List any assistance that will be provided to the evaluator by EERE. In some evaluation plans, this assistance may be detailed in the individual tasks for which it is needed.

7.3 *Schedule

The schedule should include the following milestones:

- *Project initiation meeting and due date for any minutes required for this meeting.
- Beginning and end of major data-collection activities.
- *All draft and final deliverable reports and data, including routine progress reports.

- Dates for progress meetings.
- Date for a presentation(s) of the final report, if required.
- The schedule might be presented in a Gantt chart or other type of diagram.

8. *Quality Assurance Procedure

- *Describe the quality assurance procedures that the evaluator will employ to establish confidence in the findings. These procedures should cover data collection, analysis, and reporting.
- *Specify when peer reviews will be conducted, the time allotted to these reviews (in order to remain on schedule), and who will conduct them.

9. *Organization and Management

- *Specify EERE's and the evaluator's principal contacts for the study.
- Specify the task leaders for the principal tasks.
- Specify the expected number of trips for project management if these are not specified in Task 1.
- The evaluator's staff assignments may be listed if the program manager desires it and if they will differ from the evaluator's proposal.

10. Appendices (if any)

For example:

- Successful questionnaires from related studies to serve as models for the evaluation.
- Layout of the population database that will be used as the source of a sample or census.
- Lists and contact information for program staff who will be interviewed.

Appendix 11
American Evaluation Association Ethical Principles for Evaluators

The American Evaluation Association developed the ethical principles in this Appendix to guide the professional practice of evaluators. EERE expects evaluation experts who perform general program evaluations of its programs to be governed by these principles.

American Evaluation Association
Guiding Principles for Evaluators[1]

Revisions reflected herein ratified by the AEA membership, July 2004

Preface: Assumptions Concerning Development of Principles

A. Evaluation is a profession composed of persons with varying interests, potentially encompassing but not limited to the evaluation of programs, products, personnel, policy, performance, proposals, technology, research, theory, and even of evaluation itself. These principles are broadly intended to cover all kinds of evaluation. For external evaluations of public programs, they nearly always apply. However, it is impossible to write guiding principles that neatly fit every context in which evaluators work, and some evaluators will work in contexts in which following a guideline cannot be done for good reason. The Guiding Principles are not intended to constrain such evaluators when this is the case. However, such exceptions should be made for good reason (e.g., legal prohibitions against releasing information to stakeholders), and evaluators who find themselves in such contexts should consult colleagues about how to proceed.

B. Based on differences in training, experience, and work settings, the profession of evaluation encompasses diverse perceptions about the primary purpose of evaluation. These include but are not limited to the following: bettering products, personnel, programs, organizations, governments, consumers and the public interest; contributing to informed decision making and more enlightened change; precipitating needed change; empowering all stakeholders by collecting data from them and engaging them in the evaluation process; and experiencing the excitement of new insights. Despite that diversity, the common ground is that evaluators aspire to construct and provide the best possible information that might bear on the value of whatever is being evaluated. The principles are intended to foster that primary aim.

C. The principles are intended to guide the professional practice of evaluators, and to inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators. Of course, no statement of principles can anticipate all situations that arise in the practice of evaluation. However, principles are not just guidelines for reaction when something goes wrong or when a dilemma is found. Rather, principles should proactively guide the behaviors of professionals in everyday practice.

D. The purpose of documenting guiding principles is to foster continuing development of the profession of evaluation, and the socialization of its members. The principles are meant to stimulate discussion about the proper practice and use of evaluation among members of the profession, sponsors of evaluation, and others interested in evaluation.

E. The five principles proposed in this document are not independent, but overlap in many ways. Conversely, sometimes these principles will conflict, so that evaluators will have to choose among them. At such times evaluators must use their own values and knowledge of the setting to determine the appropriate response. Whenever a course of action is unclear,

[1] The American Evaluation Association provides these Guiding Principles online at www.eval.org/Publications/GuidingPrinciples.asp.

evaluators should solicit the advice of fellow evaluators about how to resolve the problem before deciding how to proceed.

F. These principles are intended to supercede any previous work on standards, principles, or ethics adopted by AEA or its two predecessor organizations, the Evaluation Research Society and the Evaluation Network. These principles are the official position of AEA on these matters.

G. These principles are not intended to replace standards supported by evaluators or by the other disciplines in which evaluators participate.

H. Each principle is illustrated by a number of statements to amplify the meaning of the overarching principle, and to provide guidance for its application. These illustrations are not meant to include all possible applications of that principle, nor to be viewed as rules that provide the basis for sanctioning violators.

I. These principles were developed in the context of Western cultures, particularly the United States, and so may reflect the experiences of that context. The relevance of these principles may vary across other cultures, and across subcultures within the United States.

J. These principles are part of an evolving process of self-examination by the profession, and should be revisited on a regular basis. Mechanisms might include officially-sponsored reviews of principles at annual meetings, and other forums for harvesting experience with the principles and their application. On a regular basis, but at least every five years, these principles ought to be examined for possible review and revision. In order to maintain association-wide awareness and relevance, all AEA members are encouraged to participate in this process.

The Principles

A. Systematic Inquiry: Evaluators conduct systematic, data-based inquiries.

1. To ensure the accuracy and credibility of the evaluative information they produce, evaluators should adhere to the highest technical standards appropriate to the methods they use.

2. Evaluators should explore with the client the shortcomings and strengths both of the various evaluation questions and the various approaches that might be used for answering those questions.

3. Evaluators should communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret and critique their work. They should make clear the limitations of an evaluation and its results. Evaluators should discuss in a contextually appropriate way those values, assumptions, theories, methods, results, and analyses significantly affecting the interpretation of the evaluative findings. These statements apply to all aspects of the evaluation, from its initial conceptualization to the eventual use of findings.

B. Competence: Evaluators provide competent performance to stakeholders.

1. Evaluators should possess (or ensure that the evaluation team possesses) the education, abilities, skills and experience appropriate to undertake the tasks proposed in the evaluation.

2. To ensure recognition, accurate interpretation and respect for diversity, evaluators should ensure that the members of the evaluation team collectively demonstrate cultural competence. Cultural competence would be reflected in evaluators seeking awareness of their own culturally-based assumptions, their understanding of the worldviews of culturally-different participants and stakeholders in the evaluation, and the use of appropriate evaluation strategies and skills in working with culturally different groups. Diversity may be in terms of race, ethnicity, gender, religion, socio-economics, or other factors pertinent to the evaluation context.

3. Evaluators should practice within the limits of their professional training and competence, and should decline to conduct evaluations that fall substantially outside those limits. When declining the commission or request is not feasible or appropriate, evaluators should make clear any significant limitations on the evaluation that might result. Evaluators should make every effort to gain the competence directly or through the assistance of others who possess the required expertise.

4. Evaluators should continually seek to maintain and improve their competencies, in order to provide the highest level of performance in their evaluations. This continuing professional development might include formal coursework and workshops, self-study, evaluations of one's own practice, and working with other evaluators to learn from their skills and expertise.

C. Integrity/Honesty: Evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.

1. Evaluators should negotiate honestly with clients and relevant stakeholders concerning the costs, tasks to be undertaken, limitations of methodology, scope of results likely to be obtained, and uses of data resulting from a specific evaluation. It is primarily the evaluator's responsibility to initiate discussion and clarification of these matters, not the client's.

2. Before accepting an evaluation assignment, evaluators should disclose any roles or relationships they have that might pose a conflict of interest (or appearance of a conflict) with their role as an evaluator. If they proceed with the evaluation, the conflict(s) should be clearly articulated in reports of the evaluation results.

3. Evaluators should record all changes made in the originally negotiated project plans, and the reasons why the changes were made. If those changes would significantly affect the scope and likely results of the evaluation, the evaluator should inform the client and other important stakeholders in a timely fashion (barring good reason to the contrary, before proceeding with further work) of the changes and their likely impact.

4. Evaluators should be explicit about their own, their clients', and other stakeholders' interests and values concerning the conduct and outcomes of an evaluation.

5. Evaluators should not misrepresent their procedures, data or findings. Within reasonable limits, they should attempt to prevent or correct misuse of their work by others.

6. If evaluators determine that certain procedures or activities are likely to produce misleading evaluative information or conclusions, they have the responsibility to communicate their concerns and the reasons for them. If discussions with the client do not resolve these concerns, the evaluator should decline to conduct the evaluation. If declining the assignment is unfeasible or inappropriate, the evaluator should consult colleagues or relevant stakeholders about other proper ways to proceed. (Options might include discussions at a higher level, a dissenting cover letter or appendix, or refusal to sign the final document.)

7. Evaluators should disclose all sources of financial support for an evaluation, and the source of the request for the evaluation.

D. Respect for People: Evaluators respect the security, dignity and self-worth of respondents, program participants, clients, and other evaluation stakeholders.

1. Evaluators should seek a comprehensive understanding of the important contextual elements of the evaluation. Contextual factors that may influence the results of a study include geographic location, timing, political and social climate, economic conditions, and other relevant activities in progress at the same time.

2. Evaluators should abide by current professional ethics, standards, and regulations regarding risks, harms, and burdens that might befall those participating in the evaluation; regarding informed consent for participation in evaluation; and regarding informing participants and clients about the scope and limits of confidentiality.

3. Because justified negative or critical conclusions from an evaluation must be explicitly stated, evaluations sometimes produce results that harm client or stakeholder interests. Under this circumstance, evaluators should seek to maximize the benefits and reduce any unnecessary harms that might occur, provided this will not compromise the integrity of the evaluation findings. Evaluators should carefully judge when the benefits from doing the evaluation or in performing certain evaluation procedures should be foregone because of the risks or harms. To the extent possible, these issues should be anticipated during the negotiation of the evaluation.

4. Knowing that evaluations may negatively affect the interests of some stakeholders, evaluators should conduct the evaluation and communicate its results in a way that clearly respects the stakeholders' dignity and self-worth.

5. Where feasible, evaluators should attempt to foster social equity in evaluation, so that those who give to the evaluation may benefit in return. For example, evaluators should seek to ensure that those who bear the burdens of contributing data and incurring any risks do so willingly, and that they have full knowledge of and opportunity to obtain any benefits of the evaluation. Program participants should be informed that their eligibility to receive services does not hinge on their participation in the evaluation.

6. Evaluators have the responsibility to understand and respect differences among participants, such as differences in their culture, religion, gender, disability, age, sexual orientation and ethnicity, and to account for potential implications of these differences when planning, conducting, analyzing, and reporting evaluations.

E. Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

1. When planning and reporting evaluations, evaluators should include relevant perspectives and interests of the full range of stakeholders.

2. Evaluators should consider not only the immediate operations and outcomes of whatever is being evaluated, but also its broad assumptions, implications and potential side effects.

3. Freedom of information is essential in a democracy. Evaluators should allow all relevant stakeholders access to evaluative information in forms that respect people and honor promises of confidentiality. Evaluators should actively disseminate information to stakeholders as resources allow. Communications that are tailored to a given stakeholder should include all results that may bear on interests of that stakeholder and refer to any other tailored communications to other stakeholders. In all cases, evaluators should strive to present results clearly and simply so that clients and other stakeholders can easily understand the evaluation process and results.

4. Evaluators should maintain a balance between client needs and other needs. Evaluators necessarily have a special relationship with the client who funds or requests the evaluation. By virtue of that relationship, evaluators must strive to meet legitimate client needs whenever it is feasible and appropriate to do so. However, that relationship can also place evaluators in difficult dilemmas when client interests conflict with other interests, or when client interests conflict with the obligation of evaluators for systematic inquiry, competence, integrity, and respect for people. In these cases, evaluators should explicitly identify and discuss the conflicts with the client and relevant stakeholders, resolve them when possible, determine whether continued work on the evaluation is advisable if the conflicts cannot be resolved, and make clear any significant limitations on the evaluation that might result if the conflict is not resolved.

5. Evaluators have obligations that encompass the public interest and good. These obligations are especially important when evaluators are supported by publicly-generated funds; but clear threats to the public good should never be ignored in any evaluation. Because the public interest and good are rarely the same as the interests of any particular group (including those of the client or funder), evaluators will usually have to go beyond analysis of particular stakeholder interests and consider the welfare of society as a whole.

Background

In 1986, the Evaluation Network (ENet) and the Evaluation Research Society (ERS) merged to create the American Evaluation Association. ERS had previously adopted a set of standards for program evaluation (published in New Directions for Program Evaluation in 1982); and both organizations had lent support to work of other organizations about evaluation guidelines. However, none of these standards or guidelines were officially adopted by AEA, nor were any other ethics, standards, or guiding principles put into place. Over the ensuing years, the need for such guiding principles was discussed by both the AEA Board and the AEA membership. Under the presidency of David Cordray in 1992, the AEA Board appointed a temporary committee

chaired by Peter Rossi to examine whether AEA should address this matter in more detail. That committee issued a report to the AEA Board on November 4, 1992, recommending that AEA should pursue this matter further. The Board followed that recommendation, and on that date created a Task Force to develop a draft of guiding principles for evaluators. The task force members were:

William Shadish, Memphis State University (Chair)
Dianna Newman, University of Albany/SUNY
Mary Ann Scheirer, Private Practice
Chris Wye, National Academy of Public Administration

The AEA Board specifically instructed the Task Force to develop general guiding principles rather than specific standards of practice. Their report, issued in 1994, summarized the Task Force's response to the charge.

Process of Development. Task Force members reviewed relevant documents from other professional societies, and then independently prepared and circulated drafts of material for use in this report. Initial and subsequent drafts (compiled by the Task Force chair) were discussed during conference calls, with revisions occurring after each call. Progress reports were presented at every AEA board meeting during 1993. In addition, a draft of the guidelines was mailed to all AEA members in September 1993 requesting feedback; and three symposia at the 1993 AEA annual conference were used to discuss and obtain further feedback. The Task Force considered all this feedback in a December 1993 conference call, and prepared a final draft in January 1994. This draft was presented and approved for membership vote at the January 1994 AEA board meeting.

Resulting Principles. Given the diversity of interests and employment settings represented on the Task Force, it is noteworthy that Task Force members reached substantial agreement about the following five principles. The order of these principles does not imply priority among them; priority will vary by situation and evaluator role.

A. **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

B. **Competence:** Evaluators provide competent performance to stakeholders.

C. **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process.

D. **Respect for People:** Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.

E. **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Recommendation for Continued Work. The Task Force also recommended that the AEA Board establish and support a mechanism for the continued development and dissemination of the Guiding Principles, to include formal reviews at least every five years. The Principles were reviewed in 1999 through an EvalTalk survey, a panel review, and a comparison to the ethical

principles of the Canadian and Australasian Evaluation Societies. The 2000 Board affirmed this work and expanded dissemination of the Principles; however, the document was left unchanged.

Process of the 2002-2003 Review and Revision. In January 2002 the AEA Board charged its standing Ethics Committee with developing and implementing a process for reviewing the Guiding Principles that would give AEA's full membership multiple opportunities for comment. At its Spring 2002 meeting, the AEA Board approved the process, carried out during the ensuing months. It consisted of an online survey of the membership that drew 413 responses, a "Town Meeting" attended by approximately 40 members at the Evaluation 2002 Conference, and a compilation of stories about evaluators' experiences relative to ethical concerns told by AEA members and drawn from the *American Journal of Evaluation*. Detailed findings of all three sources of input were reported to the AEA Board in *A Review of AEA's Guiding Principles for Evaluators*, submitted January 18, 2003.

In 2003 the Ethics Committee continued to welcome input and specifically solicited it from AEA's Diversity Committee, Building Diversity Initiative, and Multi-Ethnic Issues Topical Interest Group. The first revision reflected the Committee's consensus response to the sum of member input throughout 2002 and 2003. It was submitted to AEA's past presidents, current board members, and the original framers of the Guiding Principles for comment. Twelve reviews were received and incorporated into a second revision, presented at the 2003 annual conference. Consensus opinions of approximately 25 members attending a Town Meeting are reflected in this, the third and final revision that was approved by the Board in February 2004 for submission to the membership for ratification. The revisions were ratified by the membership in July of 2004.

The 2002 Ethics Committee members were:

Doris Redfield, Appalachia Educational Laboratory (Chair)
Deborah Bonnet, Lumina Foundation for Education
Katherine Ryan, University of Illinois at Urbana-Champaign
Anna Madison, University of Massachusetts, Boston

In 2003 the membership was expanded for the duration of the revision process:

Deborah Bonnet, Lumina Foundation for Education (Chair)
Doris Redfield, Appalachia Educational Laboratory
Katherine Ryan, University of Illinois at Urbana-Champaign
Gail Barrington, Barrington Research Group, Inc.
Elmima Johnson, National Science Foundation

A Strong Energy Portfolio for a Strong America

Energy efficiency and clean, renewable energy will mean a stronger economy, a cleaner environment, and greater energy independence for America. Working with a wide array of state, community, industry, and university partners, the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy invests in a diverse portfolio of energy technologies.

For more information contact:
EERE Information Center
1-877-EERE-INF (1-877-3337-3463)
www.eere.energy.gov