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.