

SEP Evaluation Guidelines (Recovery Act)

It is important that the results achieved with funds provided by the American Recovery and Reinvestment (Recovery Act) be documented and assessed. These guidelines are provided to assist States in planning and conducting evaluations of their State Energy Program (SEP) Recovery Act programs. This evaluation guidance is divided into two parts. The first part is intended to guide the states' administrative and management efforts while the second part presents technical standards pertaining to the methods used to conduct program evaluations.

ADMINISTRATIVE AND MANAGEMENT STANDARDS

The following recommended evaluation administrative and management standards apply to the SEP national evaluation, and are provided for use by the States who elect to conduct their own SEP/ARRA evaluations. These standards allow evaluation efforts to be implemented using a number of research approaches, provide flexibility in determining how SEP/ARRA evaluation results reporting¹ objectives are met, and avoid the necessity for states to acquire significant new staff resources or evaluation management capabilities.

1. **Evaluation Metrics:** All projects supported by SEP/ARRA funds should be evaluated via an evaluation process that focuses on reporting metrics which reflect the principal objectives of the State Energy Program. The national evaluation will focus on the following list of metrics, and we recommend that the States focus on them as well, adding others as desired to reflect individual priorities:
 - a. Energy and demand savings
 - b. Renewable energy capacity and generation
 - c. Carbon emissions reductions
 - d. Job creation (including number, type, and duration)

Other possible metrics include, but are not limited to, economic impacts (in addition to job creation) and the adoption of new technologies.

2. **Independent Evaluations:** Programs must be evaluated independently in order to obtain reliable results. SEP Recovery Act evaluations should be conducted by independent evaluators who have no financial or management interests in the projects being evaluated. The evaluators should be independent professionals who do not benefit, or appear to benefit, from the study's findings, and the state program managers and administrators should have no influence on the findings of the study that is conducted.

¹¹ Evaluation results reporting are separate from SEP/ARRA progress reporting metrics.

3. **Attribution of Effects:** Evaluations of SEP Recovery Act-funded efforts should document the resulting effects (energy savings, renewable generation, carbon reductions and job creation) that are above and beyond the effects that would have been achieved without those funds. That is, studies should focus on net effects of the SEP Recovery Act initiatives. The effects of jointly funded initiatives, such as when SEP Recovery Act funds are combined with funds from other programs or financial offerings, will be allocated to the Recovery Act in proportion to the percentage of those funds in relation to total program or project funding.
4. **Evaluation Budgeting:** Evaluation budgets should be sufficient to ensure that reliable results are generated and reported. Typically, outcome evaluations require the allocation of between 2% and 8% of the program/project budget depending on the size and type of program/projects being evaluated. However, evaluation budgets also depend on the level of research rigor applied to those studies. For planning purposes, we recommend that states allocate 5% or less of their SEP Recovery Act funds for evaluation.
5. **Timing of the Evaluation:** Planning for an evaluation (identification of key metrics, research questions, data requirements, etc.) should begin at the same time that project activities are initiated. For many states, the services of an independent evaluator may not be immediately available upon project start-up, meaning that there may be a lag in the collection of baseline data regarding some important metrics. However, such data collection should begin as soon as possible and record-keeping on project expenditures and activities should start immediately. Evaluations should be structured to provide information to program managers as early as possible while still providing necessary rigor and reliability. It would be extremely helpful to the national SEP evaluation if State evaluations are structured so that initial study results are available within 12 months of the start of the evaluation.

TECHNICAL EVALUATION STANDARDS

The following technical standards are recommended for the evaluation studies to be performed on SEP Recovery Act-funded programs. The recommendations are presented in two sections. The first section presents general design and objectivity standards that focus on establishing objective and reliable approaches. The second section contains more detailed recommendations that are to be used within the evaluation research approaches applied to individual studies.

General Design and Objectivity Standards

1. **Study Design:** The development of the evaluation approach should be independent of project administrators and implementers and should be capable of being implemented within the evaluation budget available for the study. The independent evaluator should

work with project administrators to understand the project and its operational processes and establish an evaluation approach that is reliable and cost conscious.

2. **Study Rigor and Reliability:** The study results should be reliable. This means that the study approach must be rigorous and capable of accurately assessing impacts using the relevant SEP metrics. The studies should be designed to fit within the evaluation budget without budget overruns, and should be conducted at the highest possible level of research rigor within that budget. The evaluation community has established a number of evaluation protocols that give substantial guidance on reliable evaluation approaches. These include the National Energy Efficiency Program Impact Evaluation Guide of November 2007, the US DOE Impact Evaluation Framework For Technology Deployment Programs of July 2007, and the California Evaluation Protocols of April 2006. These documents provide guidance on establishing evaluation approaches that represent state-of-the-art evaluation approaches. There are several other protocols that can be used to guide the design and implementation of the evaluation efforts². The evaluation approach should be designed in a way that provides findings with the highest level of reliability achievable with the available research budget.
3. **Threats to Validity:** The independent evaluator should assess the various threats to validity for the study design and analytical approach and develop a study plan that minimizes those threats and reduces the associated level of uncertainty. Both the evaluation plan and the study report should identify these threats and describe how the evaluation approach minimizes threats to the validity of the study findings.
4. **Alternative Hypotheses:** To the extent possible, the study design should be developed in a way that addresses alternative hypotheses regarding how observed effects may have occurred.
5. **Ability to Replicate:** The methodological description of the study should be sufficiently detailed to allow the research design to be assessed for appropriateness by outside reviewers. The description should also be sufficiently detailed to allow the study to be replicated by other evaluation professionals.

² US EPA (1995). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program*; FEMP (2000). *Federal Energy Management Program (FEMP) M&V Guidelines: Measurement and Verification for Federal Energy Projects*. Federal Energy Management Program. September. Version 2.2, DOE/GO-102000-0960; ASHRAE (2002). *Measurement of Energy and Demand Savings, Guideline 14*. American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA.; Nexant and Lawrence Berkeley National Laboratory (2002). *Detailed Guidelines for FEMP M&V Option A*. Federal Energy Management Program.; AIS, SRC International (2001). *European Ex-post Evaluation Guidebook for DSM and EE Services Programmes*. International Energy Agency. April.; Xenergy, ADM Associates, VACom Technologies and Partnership for Resource Conservation (2001). *2001 DEER (Database for Energy Efficiency Resources) Update Study*. California Energy Commission. Study ID 3001.; Violette, Daniel (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programs*. International Energy Agency.

6. **State-of-the Art Analysis:** The study approach should, to the extent possible, use current state-of-the-art evaluation approaches that maximize the use of technical advancements and the most current analytical approaches.
7. **Unbiased Assessment:** The evaluation design, data collection efforts, analytical approach, and reporting of results should be objective and unbiased. Unsubstantiated claims or unsupported conclusions or personal points of view should be excluded and the study results should be based on objective data/information analysis.
8. **Attribution of Effects:** The study should focus on identifying the outcomes of the project in question and identify the net effects that can be attributed to the State Energy Program's implementation and support efforts.
9. **Use of Skilled Professionals:** The evaluation should employ and be led by evaluation professionals who are trained, skilled, and practiced within the area of research associated with the study being conducted.
10. **Conflict of Interest:** Evaluators must disclose any real or perceived conflicts of interest that they might have.

Study Design and Application Standards

1. **Evaluation Expertise:** The evaluation planning and implementation efforts should be directed, managed and implemented by skilled evaluation professionals experienced in the specific areas of evaluation to which they are being used to support the SEP Recovery Act evaluation efforts. Inexperienced staff should be well supervised and their work reviewed by experienced evaluation professionals for objectivity and accuracy.
2. **Study Plan:** Each evaluation should have a detailed study plan that identifies how the evaluation is to be conducted, specifying the individual tasks within the study to be completed. The study plan should also specify how data will be collected, describe processes to assure objectivity and accuracy, and identify the analysis approach to be applied for each of the four types of evaluation metrics (jobs created, carbon saved, energy generated and energy saved).
3. **Study Report:** The study report should be provided to the DOE Headquarters SEP Program Manager, with a copy to the appropriate SEP PO, and include an Executive

Summary of the results of the study. The Executive Summary should contain a table presenting:

- a. The net energy savings impacts for each year over the effective useful life of the actions attributable to the energy programs and projects supported by SEP Recovery Act funds
- b. The renewable capacity installed and the annual renewable energy generated and projected to be generated each year over the effective useful life of the installed capacity;
- c. The net tons of carbon not released into the atmosphere over the effective useful life of the projects implemented;
- d. The number and type of short term and long term full time and part time jobs generated as a result of the programs and projects supported by SEP Recovery Act funds; and
- e. The results of the SEP Recovery Act cost effectiveness test applied to the energy impacts achieved.

The study report should include the contact information for the independent evaluation contractor directing or managing the study and include their name, mailing address, telephone number and e-mail address. The selection of the evaluation contractor does not need to be approved by the US DOE/EERE/SEP manager;

4. **Sampling:** All studies that rely on sampling approaches for collecting data to drive the impact analysis objectives should, to the extent possible, use procedures that minimize bias and maximize the sample's representativeness of the targeted population. Sampling should be structured to be no less rigorous than a 90% level of precision with a confidence limit of plus or minus 10% for the key attributes on which the sample is being selected.
5. **IPMVP Field Efforts:** Field measurements of equipment baseline and post-retrofit or post installation operations should be conducted using one of the four primary data collection protocols specified in the IPMVP (International Performance Measurement and Verification Protocol). This protocol describes the types of field data collection typically used by the evaluation industry to obtain measurements needed to calculate energy impacts. This protocol describes IPMVP options A, B, C, & D for both single project end use and whole building actions. The IPMVP requires that key performance indicators that drive the estimates of program impacts should be collected via on-site metering, monitoring and verification efforts. The protocol requires measurements to be collected that represent key savings calculation indicators.
6. **Survey and Interviews:** When surveys and interviews are used to collect data from which impacts are calculated, the questions should be objective, unbiased and non-leading. Closed-ended, scaled, or quantitative response questions should be structured to

allow a full range of applicable responses. Open-ended questions should be single subject response questions that allow for a complete response. Complex questions that require a preamble to set a stage for a response consideration should be avoided to help assure that the response is objective and not guided toward a specific outcome.

7. **Cost Effectiveness Test:** The SEP Recovery Act Financial Assistance Funding Opportunity Announcement of March 12, 2009 published by the USDOE specifies that “Each state portfolio of projects funded by SEP ARRA grants should seek to achieve annual energy savings of at least 10 million source BTUs for each \$1,000 of total investment.³” This cost effectiveness test means that, on average across each state’s portfolio of programs, the energy impacts to be achieved should be no less than 10 million source BTUs⁴ per year per \$1,000 of SEP Recovery Act funds spent. These energy savings will recur each year over the effective useful life of the actions induced by the state’s portfolio. The evaluations conducted using SEP Recovery Act Funds should calculate and report the results from this test for the projects evaluated. The evaluation report should present the results of this cost effectiveness test in the Executive Summary of the report and present the calculation approach in the test in enough detail that the test can be replicated from the information presented in the evaluation report. This test is called the SEP Recovery Act Cost Test (SEP-RAC test). There are no other cost effectiveness test requirements for SEP Recovery Act project portfolios. The cost effectiveness test normally required within state regulatory environments that are focused on least cost net present value energy supplies do not apply to the SEP Recovery Act projects. DOE’s objective is to achieve deep lasting savings that provide net energy efficiency, renewable energy, carbon reductions and job impacts well into the long-term future of the United States.
8. Comments and questions relating to the above standards (both administrative and technical) should be addressed to Faith Lambert at 202-586-2319 or faith.lambert@ee.doe.gov.

³ See: Energy Savings, Section 5.7, Page 28.

⁴ Source BTU: The energy content of the fuel needed to supply the energy saved, For example, end use natural gas savings has a BTU content of about 100,000 BTUs per therm; the BTU content of electric savings will depend on the fuel source of the energy saved and the generation efficiency of the power plant to which the savings apply. A coal fired plant that is about 33% efficient would save about 10,000 BTUs per kWh saved. A savings of electricity from a hydroelectric power plant would have no BTU savings and no carbon savings because carbon fuel is not burned to provide the kWh saved.

